

Hope for best, prepare for worst

Enigmatic North Korea has government officials around the world anxious. Scientists in East Asia should be thinking about how they might be involved.

The people of East Asia are nervous. North Korean President Kim Jong-Il probably has, or will have, nuclear weapons. A recent poll carried out by a popular Japanese news programme found him the “scariest world leader” (52%), followed by United States President George Bush (34%). With negotiations between these two countries decidedly rocky, the future security of East Asian countries cannot be taken for granted.

Moreover, there is tension between the United States on one hand and South Korea and Japan on the other over dealing with North Korea. This is forcing the East Asian countries, although still largely dependent on the United States for their defence, to take a second look at their own militaries. As part of this assessment, governments in these countries, along with others such as Taiwan, will be reconsidering their nuclear posture.

Scientists must be involved with these debates, both domestically and regionally. In the past, much of the tough thinking concerning the military application of science was the preserve of scientists in the United States, the Soviet Union, and other key players in the cold-war arena. Since the Manhattan Project, American scientists have sought, rightly, to play a central role in helping to shape decisions about how their knowledge should be used.

This effort has not always been successful, but it has been the least that could be expected ethically from the people whose technical expertise made nuclear weapons possible. (Many of those scientists who did not adequately consider the larger picture early in the development of nuclear weapons admitted regret in the following years.)

The consciousness concerning such issues in East Asia appears to be minimal (see pages 110–111). Scientists are adopting the attitude that war is something for the government to deal with. This attitude reflects their ideal for their lives in research: don’t let political considerations skew your objectivity; focus on your research.

A researcher’s career, so this argument runs, is defined by following pure curiosity, not by getting hitched to some political agenda. In Japan especially, researchers try to follow this ideal — so much so that would-be entrepreneurs have trouble getting university professors involved in start-up companies.

No one wishes to encourage researchers who, for example, take advantage of tensions in diplomacy to seek increased funds for their research project. But researchers in the West and even India and Pakistan seem to be more outspoken in promoting the military role of scientists. In East Asia, in contrast, they seem to be happy to stay in the background, to focus on their research.

But just because they do not seek active involvement does not mean that they and their research will not be enlisted. Military preparations require science-based research and development, from the analysis of an enemy’s capabilities to the design and production of weapons.

If researchers are willing to leave decisions considering the military application of science to politicians, they will all too possibly become involved in ways that they could never have anticipated. Like it or not, scientific research — especially that in military-related fields — plays a political role. If leading researchers are not actively engaged in defining this role, those much less qualified will do so. ■

Berlin’s university crisis

German universities must take drastic decisions, to pursue a more flexible approach to education and research.

It is a sad twist of fate that Berlin’s venerable Humboldt University is associated with the most alarming shock to Germany’s academic system since the war. Responding to announced cuts of up to €600 million (US\$680 million) in its 2006–09 budget, the university has said that it will not admit students for next winter term as it cannot guarantee proper scientific training and reasonably quick completion of studies.

Alexander von Humboldt, the nineteenth-century Berlin-born naturalist and geographer, was the founder of the modern German higher-education system. Its so far unshakeable principle, the unity of teaching and research, made Germany a cradle of modern science in the first half of the twentieth century. The success of science in Europe, the United States and Japan can be traced back to Humboldt’s university ideals.

It is not difficult to see why, while Harvard and Oxford flourish, the Berlin original is on crutches. Berlin is still struggling with the immense costs of German reunification. Not only its three universities but also its opera houses, theatres, kindergartens and welfare facilities are facing major budget crises.

But the Humboldt University’s unprecedented move — an act of despair committed on shaky legal ground — illustrates a fundamental

problem of the higher-education system throughout Germany. Only a small minority of the 200,000 students who enrol each year at its 100 or so research universities is seeking a full scientific education. The majority would be satisfied with a solid academic training that qualifies them for a decent job outside research. But German universities aim at making researchers of them all. And they offer this for free: tuition fees are unknown in Germany.

This has much to do with deeply rooted anti-elitism, based on the respectable belief that everybody should have the same educational opportunities. Only in the past few years has hyper-egalitarianism rightly begun to be questioned. The recent introduction of bachelor degrees and fees for long-term students is surely only the start of change. The allocation of scarce resources must be rethought, and not only in Berlin, in order to keep the system competitive and viable.

The current merger of Berlin’s medical faculties is a step in the right direction. Germany-wide, excellence in research needs selectivity without sacrificing training opportunities for as many as possible. Not every university need be a research hot-spot, and not every student needs to be trained by world-class scientists. If further down the road the Humboldt University becomes a centre of excellence in a more diversified university landscape, so much the better. ■





On guard
Vigil for West Nile virus calls on US chicken flocks
p104



Patent power
Geneticists face prospect of paying to use junk DNA
p105



Targeted effort
The search for a therapeutic AIDS vaccine
p107



Strange finding
Enigmatic quark discovery charms physicists
p108

France purges space programme in bid to survive budget crisis

Declan Butler, Paris

In a dramatic bid to address its long-running financial problems, the CNES, France's national space agency, has cut or frozen 10 of its 44 missions — including an ambitious effort to send four landers to Mars.

The agency has struggled over the past five years with falling government funding. Last year, it overspent its €1.3-billion (US\$1.5-billion) budget by €90 million.

The rescue plan was announced on 30 April at a press briefing in Paris by Yannick d'Escatha, a former head of the French Atomic Energy Commission who was brought in by president Jacques Chirac to reform the space agency. "We are in a crisis situation," d'Escatha said. "We cannot do everything."

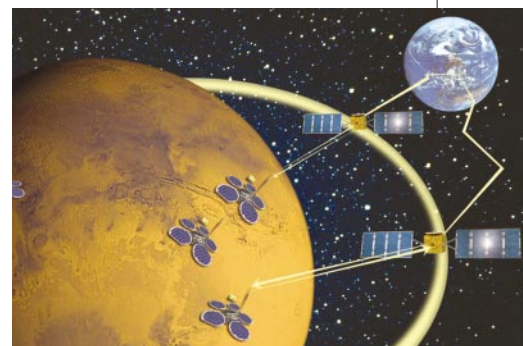
The largest victim of the cuts is NetLander, a French-led mission to Mars. The mission would have dropped four landers on the planet, carrying sensors to study its geophysics, search for water, and record atmospheric conditions. The CNES planned to spend €100 million on the project, which also involved Finland, Germany, Belgium

and Switzerland. In March, NASA shelved plans to contribute \$35 million to the project.

In addition, the CNES will withdraw its stake in the NASA-led Gamma-ray Large Area Space Telescope (GLAST). Also put on hold are four experiments planned for the International Space Station.

NetLander's collapse marks the demise of an ambitious Franco-American programme of Mars exploration, announced with considerable fanfare in October 2000. The CNES will now support Mars-exploration missions only under the auspices of the European Space Agency (ESA), says d'Escatha.

Space officials on both sides of the Atlantic deny that the loss of these missions is linked to the current political rift between the United States and France. But researchers say that tight budgets and political tensions are damaging international programmes. "Five years ago there was a strong desire by everybody from the middle levels of NASA up to the White House for increasing international cooperation," says Bruce Banerdt, principal investigator for NetLander at NASA's Jet Propulsion



In the red: falling funds have led the French space agency to cancel its NetLander Mars mission.

Laboratory in Pasadena, California. "There is certainly a different atmosphere now."

France's other bilateral space-exploration collaborators have also been hit. The CNES says that it will renegotiate its partnership with India to launch the Megha-Tropiques satellite, to monitor the atmosphere in the tropics. The cuts will spell doom for a proposed Franco-Brazilian atmospheric satellite, and a ground station for the Japanese satellite Alos.

Among missions that have survived are the space telescope Corot, the ocean-surveillance mission Jason-2, and Pléiades, a group of high-resolution Earth-observation satellites.

Researchers are worried, however, that the cuts amount to a downgrading of science at the CNES to pay for other priorities, including supporting the satellite and telecommunications industries, military space programmes and the assurance of Europe's independent access to space aboard the Ariane rocket. The agency's financial health has not been helped by the crisis in the commercial launch market due to a drop in the number of telecoms satellites, and the recent crash of Ariane V (see *Nature* 420, 723; 2002).

D'Escatha has written to NetLander's scientists to suggest that the mission might be saved by incorporating it into ESA's space-science programme, or its Aurora planetary-exploration programme. But David Southwood, ESA's director of science, says that the former option is unlikely within the current budget. A decision on Aurora is unlikely to be taken before next year.

US Army joins hunt for SARS drug

Alison Abbott

Researchers at the National Institutes of Health and the US Army have joined forces in a systematic screening programme to find drug candidates to combat severe acute respiratory syndrome (SARS).

The collaboration, which is already under way, was announced at last week's International Conference on Antiviral Research in Savannah, Georgia.

The National Institute of Allergy and Infectious Diseases is collating as many compounds as it can gather that are licensed or in development as antiviral drugs. The US Army Medical Research Institute of Infectious Diseases (USAMRIID) at Fort Detrick, Maryland, will then screen them for activity against the coronavirus thought to cause SARS. In addition, the researchers are screening a commercial library of 400,000

chemicals that are yet to be licensed.

"We hope to run through hundreds of such compounds — or even a thousand — very quickly, starting with those already licensed, and working back along the development pipeline," says Robert Baker, a USAMRIID virologist. With luck, he says, the programme could identify agents that will be available for treating SARS within months.

In Europe, meanwhile, antiviral expert Erik De Clercq of the Catholic University of Leuven, Belgium, is taking a different tack. He says he will first identify key points in the biosynthesis of SARS viral RNA that could be targeted by drugs. His group will then evaluate selected compounds from its own chemical library, as well as those submitted by colleagues at other centres. "It's a more 'rational' approach," he says. ■

US fails to quantify threat of West Nile virus

Hannah Hoag, Washington

Public-health authorities in North America are gearing up for the return and probable spread of West Nile virus, which last year claimed the lives of more than 200 people.

But despite a national effort involving several federal agencies, including a \$40-million investment in research and surveillance by the Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, officials still have no reliable estimate of the likely scale of this year's outbreak.

West Nile virus is carried by birds and is mainly spread to humans by mosquitoes. It is a member of the flavivirus family, which includes yellow fever and St Louis encephalitis, but in humans it is not usually life-threatening, although in about 1% of cases it causes severe meningitis and encephalitis.

Since it was detected in New York state in 1999, the virus has surged westwards. Last year it infected at least 4,000 people in the United States, killing 284, and public-health officials are predicting that the virus will spread to cover the whole country this year.

But virologists also expect both humans and animal host populations in the United States to begin developing immunity to the virus, as has happened in other parts of the world where the virus is already endemic. However, there is so far no reliable estimate of how long this immunity will take to build up, leaving researchers unsure of the likely death toll for this year.

"We should be prepared for an outbreak just as large, if not larger than last year's," says Lyle Petersen, of the CDC's division for vector-borne diseases. "There is nothing to prevent it from happening again."

US researchers have identified 36 species of mosquito that can carry and transmit the virus to humans, although three species of *Culex* mosquito seem to be responsible for most US cases.

Last year, 21 people were infected through blood transfusions, and four became infected after receiving organ transplants from infected donors. Two companies are developing screening tests in an effort to prevent further human-to-human transmission of the virus. The tests are expected to be ready by 1 July.

But experts are criticizing the US response to the virus, describing current efforts as patchy. Because of funding cuts, some state health departments have dispensed with bird surveillance programmes, which track and identify the incidence of the virus in birds, on the premise that the virus has already become endemic.

In California, where the virus has yet to arrive and where the relatively high number of outdoor workers, such as farm labourers,



Blood is taken from a Californian chicken as part of an early warning screen for West Nile virus.

could heighten health risks, officials are preparing for the worst. "We fully expect that West Nile virus will be introduced and become established this summer," says Vicki Kramer, chief of the vector-borne disease section of the state's health department. California has established a network

of 200 chicken flocks which will be monitored by health officials for the arrival of the virus.

University researchers and biotechnology companies, meanwhile, are trying to develop a vaccine. Drug firm Acambis in Cambridge, Massachusetts, for example, this summer plans to begin human safety trials of a live vaccine, based on an established vaccine against yellow fever. A DNA vaccine — derived from the virus's RNA — and one derived from its protein coat are also being developed.

Other researchers are looking at how to improve control of the mosquitoes. Researchers at the US Agricultural Research Service in Gainesville, Florida, have identified a baculovirus that selectively infects and kills the *Culex* mosquito. But the team has yet to find a commercial partner to develop the idea into an insecticide that could be approved and used. ■

Los Alamos contract open to bids

Geoff Brumfiel, Washington

Researchers at Los Alamos National Laboratory in New Mexico are facing an uncertain future after the US Department of Energy (DOE) announced that it will accept bids from new contractors to run the lab.

For the whole of its 60-year history, the nuclear-weapons laboratory has been managed by the University of California.

Scientists credit the university with fostering an atmosphere of academic rigour and intellectual openness that they fear could disappear under management by a corporation or other institution. "There's great concern that the intellectual capability of this lab could just dissolve," says William Priedhorsky, chief scientist in Los Alamos' nonproliferation division.

The DOE says that its decision, announced on 30 April, is a response to recent fraud cases involving lab employees, and to inventory audits that highlighted missing or misplaced equipment (see *Nature* 421, 99–100; 2003). "The problems at the lab represent a systemic management failure," Linton Brooks, head of the National Nuclear Security Administration — the DOE agency that oversees the lab — told a congressional committee. The department will seek competitive bids to run the lab from September 2005, when the university's existing contract expires.

But some observers charge that the decision is driven in part by the Bush administration's desire to make trouble for Gray Davis, the Democrat governor

of California. Organizations close to the administration, including the University of Texas and the engineering firm Bechtel, are thought to be top candidates for running the lab. "This is ultimately a political decision," says one scientist working at the laboratory. "This administration doesn't have much love for the state of California."

Management issues at Los Alamos have also raised the ire of Congress, which has been growing impatient with the lab since Wen Ho Lee, a mechanical engineer there, was accused of passing weapons secrets to a foreign power in 1999. And now even the lab's most ardent congressional supporters, such as Senator Pete Domenici (Republican, New Mexico), are conceding that other institutions should be allowed to compete for the management contract.

Richard Atkinson, the University of California's president, says that he is unsure whether the university will compete for the contract. But he adds that if it does, it will be without fanfare or the millions of dollars usually expended to win such a role.

Los Alamos officials say that they have already begun to feel the effects of the DOE's decision. Retirement applications have almost doubled compared with the same time last year, and recruitment is also being adversely affected. "It's foolish to throw away a contractor that is one of the best publicly run universities in the world," says Albert Migliori, a physicist and 30-year veteran of the laboratory. ■

Geneticists question fees for use of patented 'junk' DNA

Carina Dennis, Sydney

Most geneticists have never heard of Malcolm Simons. But they could get to hear about him pretty soon when they're asked to pay for use of non-coding DNA — sometimes known as 'junk' DNA — on which the New Zealand immunologist has won wide-ranging global patents.

Genetic Technologies (GTG), the Australian company that now holds the rights to the patents, is starting to assert these rights in universities. And researchers could shortly need a licence from the company to use any non-coding sequence in genetic analyses of any species in their research.

"We have contacted academic research groups in Australia, New Zealand, the United States, Japan and Europe," says Mervyn Jacobson, chairman of GTG, who says that the company is in the final stages of negotiations with three universities in Australia and one in the United States.

Some academics — used to the fact that most patent-holders don't ask for license fees from basic researchers, at least until a researcher tries to make money through commercial applications — are not exactly thrilled by GTG's plans.

"I feel outraged," says Joe Sambrook, a molecular biologist at the Peter MacCallum Cancer Institute in Melbourne, Australia, and author of a well-known technical manual for molecular biologists. "I think that asking licence fees from academic researchers can only inhibit research. If they do it, other people will do it — and it has not been a common practice."

But under most nations' laws, patent-holders are perfectly at liberty to ask for licences, often in return for fees from users, including basic researchers. "Whether it is right or wrong, I don't know, but that's the law," says Deon Venter, a pathologist at the University of Melbourne who was recently appointed to oversee GTG's own programme to perform genetic tests for patients' susceptibility to, among other conditions, breast cancer.

Simons first cottoned on to the value of non-coding DNA some 15 years ago, while studying the immune system's genes. Afterwards, he successfully applied for several patents involving access to the information that is embedded in the non-coding DNA of all species.

Melbourne-based GTG has already amassed millions of dollars' worth of licensing deals from drug companies, and is now turning its attention to universities. "Researchers have nothing to be frightened of — it's not going to be financially burden-



Money-spinner: patents on junk DNA could cost university researchers dear.

some for them," argues Jacobson. "We are negotiating several at the moment for a thousand dollars — which doesn't even cover our legal costs in producing the document. We are not intending to be aggressive or hostile, or to stifle research."

It isn't clear if universities will challenge the charges, or bite the bullet and pay up. "We are very concerned about the patents," says Debra Graves, chief executive of the Royal College of Pathologists of Australasia, based in New South Wales. Graves is coordinating a submission from researchers to the Australian government on the matter. Several academics have said that their universities will seek legal advice before deciding whether a challenge is worthwhile.

GTG has already attracted controversy over its licensing deal with Myriad Genetics of Salt Lake City, Utah, which gives GTG exclusive rights to Myriad's genetic-susceptibility tests, including those for breast cancer, in Australia and New Zealand. "We plan to be the leading genetic-testing facility in the region," says Jacobson.

Meanwhile, Simons, who retired from GTG in 2000, has little to show for his foray into the DNA goldmine. Despite being seriously ill with cancer, he has lost none of his interest in genetics, claiming that current hunts for complex disease genes are off the mark. "I'd like to set them straight in the time I have left," he says.

Simons' life will soon be the subject of an Australian Broadcasting Corporation television documentary.

Physicist takes the reins at Santa Fe complexity centre

Geoff Brumfiel, Washington

For almost two decades, scientists at the Santa Fe Institute (SFI) in New Mexico have been struggling to describe some of the world's most complicated systems. But simplicity is the goal of their new boss, a straight-talking nuclear physicist named Robert Eisenstein.

Eisenstein is a seasoned administrator who ran the physics directorate at the National Science Foundation (NSF) from 1997 to 2001. He was then seconded from the NSF to CERN, the European particle-physics lab, where he has been helping to install a 7,000-tonne detector called ATLAS.

Last month, he was named president of the SFI, which specializes in the study of complexity — an approach to the modelling of everything from the pattern of spots on a cheetah to the movements of the stock market.

The SFI has always relied on informal processes to keep it running and to bring in new talent, says Stuart Kauffman, a biophysicist and member of the SFI's science board. When Eisenstein takes over next month, he may take a more structured management approach.

He has a knack for grasping the big picture and "not getting overwhelmed by the details", says Joseph Deahmer, current head of the NSF's physics division.

The SFI was set up in 1984 by an iconoclastic group of researchers from nearby Los Alamos National Laboratory, who wanted to bring a new approach to science. They sought to apply their ideas about complexity to hard-to-quantify topics, such as cellular behaviour and evolution, and even the social sciences.

"We're all very aware of the fact that that institutions tend to lose their edge over time," notes Erica Jen, a mathematician and research professor at the SFI. She says that many see Eisenstein's appointment as a chance to

boost the recruitment of fresh talent into the institute.

Eisenstein thinks his track record at the NSF can help.

"I saw an enormous variety of different kinds of science come across my desk," he says. "And I learned a lot by watching what was happening in those programmes."



Eisenstein: straight-talking attitude.

Homeland science chief wants quick fixes

Erika Check, Washington

In his first few weeks as undersecretary for science and technology at the US Department of Homeland Security, Charles McQueary has seen proposals for countless homeland-security devices, ideas for tamper-proof manhole covers and an anti-anthrax spray, among others. But so far he hasn't awarded so much as a cent to any of these pitches.

Although he was sworn in on 9 April, McQueary is still waiting for Congress to give him his first \$30 million, which he will spend on an initial round of grants to support the directorate's technology priorities. Until then, he is spending much of his time fielding queries from businesses, universities and inventors who want a share of the \$800 million that his directorate has requested from Congress for 2004.

The early signs are that bench scientists won't see much of the money. McQueary admits that his main priority will be to find existing devices and bring them rapidly into use. For the time being at least, the focus is on applied, rather than basic, science. "It's very important that we have an understanding about what kinds of scientific solutions exist today, so we can bring those to bear quickly," he says.

McQueary, a 63-year-old mechanical engineer, came out of retirement to head the new directorate. He spent much of his career running research and development programmes at defence contractor General Dynamics, and at AT&T's Bell Laboratories in Murray Hill, New Jersey.



Charles McQueary: more interested in speedy solutions than in blue-sky research.

He is currently building up a staff of about 150 to run the science and technology directorate from the Department of Homeland Security's temporary headquarters, at a Navy complex in an opulent residential area of northwest Washington DC. The complex's wide, grassy lawns and red-brick buildings make it look like a university campus, but the number of officials strolling about in uniform — as well as the tight security at the front gate — give it the feel of a military base. The whole

department is expected to move to new headquarters before too long.

Already, McQueary's actions are being watched assiduously by university officials, hopeful contractors and those in Congress who pushed in the first place for a Homeland Security directorate dedicated to science and technology. "We'd like to get more information from the administration on what kind of research agenda they intend to pursue and why," says David Goldston, chief of staff on the House Science Committee, which advocated setting up the directorate.

One part of that agenda that will be implemented this year, McQueary says, is the creation of university-based research centres in five priority areas: radiological, chemical, biological, nuclear and high-explosive threats.

The department will also organize its own Homeland Security Institute, probably led by staff at an existing federal institution such as the Lawrence Livermore National Laboratory in California. Next year, it is also expected to award \$350 million in new funding for critical areas such as border security, through a Homeland Security Advanced Research Projects Agency.

McQueary's focus on immediate solutions, rather than blue-sky ideas, may derive from the fact that he thinks about the terrorist attacks of 11 September almost every day. "It's easy to look back and say we could have done this or that, but the fact is we weren't thinking that way," McQueary says. "The challenge is to prevent the next thing that's comparable."

Dying cells dragged screaming under the microscope

Catherine Zandonella, New York

An eerie high-pitched song fills the air in a physics laboratory in Los Angeles. It is the



"THE SCREAM" BIRCH AFTER MUNCH

scream of a yeast cell as it withers in a pool of alcohol, and it just may proclaim a useful new technique for cell biologists.

The master of this cellular torture chamber is physical chemist James Gimzewski of the University of California, Los Angeles. Best known for his work on manipulating single molecules during his career at IBM's Zurich research laboratory, Gimzewski has of late turned his nanoknowhow towards exploring single cells. He is trying to develop a new technique for studying the cell by listening to the sound that its membrane makes.

Gimzewski and graduate student Andrew Pelling have been looking closely at how the outer membranes of yeast cells vibrate, depending on the condition of the cells. Now they're teaming up with cancer pathologist Michael Teitell of the UCLA

School of Medicine to establish whether this signature song could be used to monitor the health of a cell in response to both external stresses and internal ones, such as gene mutations.

To obtain such sensitive measures of the cell membrane, which moves only about two billionths of a metre with each vibration, Gimzewski and Pelling use the nanotechnologist's favourite tool — the atomic force microscope (AFM). The AFM's tip can be scanned over a surface to create a precise map of its topography. "We use the AFM as a type of stethoscope, to follow how cell membranes move," says Gimzewski, who has christened the method 'sonocytology'.

The researchers find that strains of yeast emit slightly different sound patterns. When amplified, they sound a bit like rhythmic

AIDS researchers seek criteria for vaccines

Erika Check, Washington

Scientists have been seeking an AIDS vaccine for two decades now, with precious few signs of success. But while their search has focused primarily on a vaccine that would prevent the disease, some vaccine researchers have been looking for ways to augment existing drug treatments for AIDS.

Just as they find preliminary evidence that such a vaccine might work, however, researchers investigating such therapeutic vaccines are running into new difficulties.

Advocates of therapeutic vaccines complain, in particular, that regulators are stalling research in this field by refusing to set clear expectations for what such a vaccine needs to do. Without clear targets, scientists and companies say that they are reluctant to set up large, costly clinical trials.

But government officials counter that researchers are demanding too much guidance too soon — before they've even shown that such a vaccine could work. "I think there's been a hold-up on both sides," says Ben Cheng of the government-sponsored Forum for Collaborative HIV Research, which hosted a workshop on therapeutic vaccines in Washington DC on 23 and 24 April.

The idea behind therapeutic vaccines is to prolong the useful life of anti-retroviral therapies, which allow many patients to survive AIDS but are prone to failure over time, as the virus evolves resistance to the drugs. A few studies have hinted that a therapeutic vaccine could help to solve this problem by allowing patients to take short breaks from treatment.

In one such study, funded by the ANRS,



At the sharp end: researchers seek therapeutic vaccines to treat AIDS, as well as preventative ones.

the French national AIDS-research agency, Brigitte Autran and Christine Katlama of the Hôpital Pitié-Salpêtrière in Paris found that a canarypox vaccine let some patients take slightly longer breaks from drugs than others (see *Nature* **422**, 650; 2003). And earlier this year, Yves Levy of the ANRS reported that patients who took a vaccine and a biological booster suppressed the virus better than those who did not take a vaccine.

The question now is what regulatory agencies will do with this evidence. A preventative AIDS vaccine succeeds when it stops people becoming infected with the virus. But a therapeutic vaccine's success could take many forms: reducing the number

of HIV particles in a patient's body, for example, or allowing the patient to stay off drugs for a certain period of time.

So far, the US Food and Drug Administration (FDA) and other agencies have not told scientists which of these criteria it will consider as proof of success. But at the workshop, Carol Weiss, a scientist at the FDA's Office of Vaccines Research and Review, gave some indications. "The treatment needs to change a patient's outcome in a clinically meaningful way," Weiss said, adding that the change must be "durable". How durable, she said, has not yet been determined, but she hinted that any approved vaccine would have to work for at least six months. Weiss also said that an FDA advisory committee would probably have to approve the first large trial of any therapeutic vaccine.

That statement will not satisfy everyone. But Dan Kuritzkes, director of AIDS research at the Partners AIDS Research Center at Massachusetts General Hospital, says that the FDA's comments will help scientists. "I think we came out with a much clearer sense of what the potential indications might be for therapeutic vaccines," such as a lower viral load or an extended time off treatment, Kuritzkes says. He adds that this is especially important for commercial vaccine development: "Industry needs to have a sense that if they get to the end of the road, the agencies will have similar expectations about what is reasonable."

Alan Landay, an immunologist at Rush-Presbyterian-St Luke's Medical Center in Chicago, says that the workshop marked a milestone for therapeutic vaccines in at least one sense: by the end of it, much of the tension between regulators and scientists had dissipated. "Everyone in the room saw that others were facing the same sorts of questions they had," Landay says. "Now we all know we're not alone in this quest."

breathing. When researchers dribble some isopropyl alcohol onto one of the yeast cells, the pitch increases sharply, as if the cell is emitting a shrill cry. A dead cell produces the kind of static you might hear when tuning a radio.

Researchers have previously used an AFM to detect the pulsing surface of cardiac cells, which 'beat' even when removed from the body (S. G. Shroff, D. R. Saner and R. Lal, *Am. J. Physiol. Cell Physiol.* **269**, C286–C292; 1995). But Gimzewski thinks that he's the first to listen to the membrane resonance of other cell types.

Gimzewski — a high-flyer at IBM who has published frequently in major scientific journals, including *Nature* — hasn't yet published anything on his idea. But he has discussed it with the *LA Weekly*, an alternative newspaper in Los Angeles, and

he plans to demonstrate it this autumn at the Los Angeles County Museum of Art, in a nanotechnology exhibition held jointly with an artist with whom he has collaborated before, Victoria Vesna. He says that this approach fits in with his goal of taking science to a wider audience.

Sceptics point out that the technique may not work for mammalian cells, which have much softer membranes than yeast cells. But Jan Hoh, a physiologist at Johns Hopkins School of Medicine, says that he is intrigued by the concept of sonocytology. "If there are frequencies in the kilohertz range that truly are diagnostic to something going on in the cell, that would be interesting," says Hoh. He cautions that treating cells with alcohol is too crude to reveal much, however.

Shuttle's wonder worms make it back to NASA in one piece

Washington Live nematodes recovered from the debris of the space shuttle Columbia have finally made their way back to the scientists who sent the worms — or rather, their ancestors, as they only live for a week or two — on the ill-fated voyage.

Between 1,000 and 10,000 *Caenorhabditis elegans* worms were found in petri-dishes packaged inside one of the shuttle's small mid-deck lockers, according to Nate Szewczyk of NASA's Ames Research Center in Moffett Field, California, a member of the team that was testing a new kind of synthetic medium for nourishing the nematodes during spaceflight.

C. elegans is a handy model organism for space biology studies, and has already flown on several shuttle flights. Szewczyk says that once the Ames researchers realized the equipment housing the worms had made it through re-entry, they were not surprised that the animals survived the violent forces that broke Columbia apart. Nematodes have endured brief exposures to accelerations 100,000 times the force of gravity.

Mathematics institute appoints new number one

Washington The Clay Mathematics Institute, which has been rudderless since November when its president was ousted and two board members resigned, now seems to be getting back on track. Its new president will be James Carlson, a mathematician at the University of Utah in Salt Lake City.

Established in Cambridge, Massachusetts, in 1998 by financier Landon Clay, the institute has been embraced by mathematicians. It awards \$3 million of research grants each year, and has raised the profile of pure maths by offering prizes of \$1 million for anyone who solves each of seven famous problems.

Mathematicians were concerned about the sudden departure of Harvard mathematical physicist Arthur Jaffe and the subsequent resignations of two other members of the institute's four-strong scientific advisory board (see *Nature* 421, 303; 2003). The board has now been reconstituted, and Carlson will take the reins in August.

Cancer body gets tough on conflicts

San Diego The 20,000-member American Society of Clinical Oncology (ASCO) last week adopted one of the toughest conflict-of-interest policies of any scientific body.

Research leaders seeking to present studies at ASCO conferences or to publish articles in the society's journals must now

disclose any financial support from a project sponsor in excess of \$100, including gifts and travel expenses. In comparison, the US National Institutes of Health only requires researchers who receive grants to disclose to their institutions payments or stock in excess of \$10,000 a year from a study sponsor.

ASCO president Paul Bunn, an oncologist who directs the University of Colorado's Cancer Center in Denver, explains that a plethora of conflict-of-interest cases prompted the rewriting of a previous policy that dates from 1996. "We really felt a need to tighten up," says Bunn. The new policy will come fully into effect in a year's time.

♦ www.asco.org

Howard Hughes steams ahead with new campus

Washington Despite financial worries, the Howard Hughes Medical Institute (HHMI), based in Chevy Chase, Maryland, is pushing ahead with an ambitious new research campus in rural Virginia. The institute's board of trustees is expected this week to name HHMI vice-president Gerald Rubin as the first head of the new project, shortly after a groundbreaking ceremony on 5 May.

HHMI officials say that the campus, called Janelia Farm, will cost around \$500 million. It will house up to 300 investigators, as well as visiting scientists, fostering projects that will yield new research tools, says HHMI president Thomas Cech.

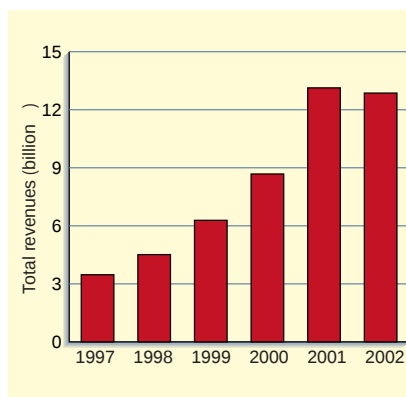
So far, Janelia Farm has escaped cuts that the HHMI announced last year, which include trimmed budgets for the 300 investigators it funds at universities around the United States. "We need to respond to new opportunities," says Cech. In November, the institute said that the continuing economic downturn had slashed its endowment from \$13 billion to \$10 billion over the preceding three years.

Quark 'compound' has them all guessing

San Francisco A new composite subatomic particle has been uncovered by physicists at the Stanford Linear Accelerator Center (SLAC) in California — but no one is sure exactly what it is.

Given the catchy name of Ds (2317), it seems to be an unusual configuration of quarks — the fundamental building blocks from which protons and neutrons are made. Ds (2317) could be an anti-strange quark orbiting a charm quark, or perhaps a 'quark molecule' made of four quarks. But whatever it is, researchers hope the particle will deepen their understanding of the strong nuclear force, which binds quarks together.

Physicist Marcello Giorgi of the University of Pisa in Italy and his colleagues stumbled on Ds (2317) while sifting through years' worth of data gathered by the BaBar detector at SLAC (pictured), which looks at the high-energy collisions between



Over the hill? Last year was the first since 1997 in which the biotech revenues in Europe fell.

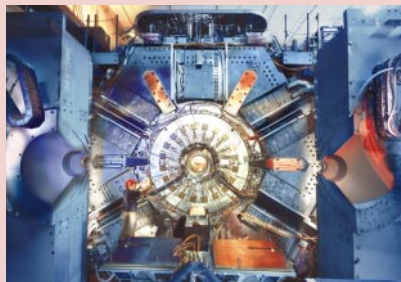
Biotechnology industry plateaus out in Europe

London Europe's biotechnology industry has stalled, according to a report from consultants Ernst & Young. The number of biotech companies on the continent, which has grown at 10–20% each year since 1997, remained almost static at around 1,900 in 2002. Revenues dropped by 2%, and the workforce decreased in size by 6%.

Some sectors have suffered more than others from the current economic slump. Genomics companies have been the hardest hit, whereas more established areas of the industry, such as vaccine development, have been less affected. The report does not cover large pharmaceutical companies.

Glenn Crocker, biotechnology adviser at Ernst & Young in Cambridge, UK, and the report's lead author, says that the industry faced similar problems during the 1990s and recovered from them. But he warns that some lessons, such as the need to base strategies on revenues generated in the early years of a business, have not yet been learned.

♦ www.ey.com/uk/healthsciences



electrons and their antimatter counterparts. The particle was found among relatively low-energy debris from these collisions. It is unusual in that its mass is lower than expected, and is well defined — typically, the mass of such composite particles is made fuzzy by quantum uncertainties.

♦ www.slac.stanford.edu

Asia's nuclear family

If North Korea has nuclear weapons, its neighbours may want to develop their own. Geoff Brumfiel and David Cyranoski ask whether Japanese and South Korean scientists would answer a call to nuclear arms.

Hirota Sugawara has an unlikely proposal. One of Japan's most prominent high-energy physicists, he has spent his working life studying sub-atomic particles. Now a visiting researcher at the University of Hawaii, Sugawara wants to use a beam of one such particle — the ghost-like neutrino, which passes through most matter — to root out and destroy stocks of hidden nuclear weapons. "It's rather futuristic," he admits.

Sugawara says the beam will interact with any plutonium it encounters, melting or even igniting the weapon. The proposal may never be tested, but the fact that Sugawara felt the need to suggest building a device that could detonate an enemy state's nuclear weapons shows how concern over nuclear proliferation is affecting researchers in Japan.

Sugawara doesn't mention a nation on which he would use the beam, but North Korea is one obvious target. The country could develop, or may already have developed, a nuclear device. This could leave many scientists in Southeast Asia facing an unwelcome decision: knowing that a feared and unpredictable neighbour is developing a frightening military advantage, should they support nuclear-weapons programmes in their own countries, or resist such proliferation?

North Korea hit the headlines last month, when its officials claimed to have developed nuclear weapons, but tension in the region has been high since late last year. Under a 1994 agreement, the country halted development of its nuclear arsenal in exchange for funding from the United States, Japan and South Korea for energy projects. The deal remained intact until last October, when the United States announced it had evidence that North Korea had gained new nuclear-weapons technology.

Since then, North Korea has appeared determined to let the world know of its nuclear plans. In December, it kicked out the inspectors monitoring its mothballed nuclear plants. A month later, it pulled out of the Nuclear Non-Proliferation Treaty and, a month after that, restarted its nuclear facilities. And in February and March this year, it test-fired two unarmed missiles into the Sea of Japan.

North Korea has since said that it does indeed have nuclear weapons, but has also proposed a deal to end its nuclear programme if the United States pledged not to attack. Given North Korea's history of deliberate obfuscation, observers warn against reading too much into either statement.



Arms race: many fiercely ideological North Koreans see nuclear weapons as vital for their nation.

Experts agree, however, that North Korea has the potential to back up its sabre-rattling. A hundred kilometres north of the capital Pyongyang, the town of Yongbyon is home to a 5-megawatt nuclear reactor and a plant for reprocessing its spent fuel. The reactor runs on uranium, producing waste from which the reprocessing plant can extract plutonium — which could then be used in a nuclear weapon.

True believers

Only a handful of outside scientists and technical experts have been to Yongbyon. Robert Alvarez, an expert in nuclear programmes at the Institute for Policy Studies in Washington DC, visited in 1994 and 1995 with teams from the US Department of Energy. He estimates that a few hundred researchers work there — most trained at Soviet institutions, although the younger ones studied at Pyongyang's Kim Il Sung University. They are passionate followers of North Korea's philosophy of independence from foreign powers, and believe nuclear weapons will ensure their nation's security. "This is their life's work," says Alvarez. "These guys are fiercely dedicated."

Their dedication has endured in harsh and dangerous working conditions. Alvarez recalls lecture rooms that lacked heat and electricity. What's more, the nuclear facilities at Yongbyon are based on early British and US designs, published in the open literature, which are now generally considered unsafe and unreliable.

The reactor is an early version of a 'Magnox'

facility, the first of which was opened at Calder Hall in Cumbria, northwest England, in 1956. Improved versions are still used in Britain, but the original plants are considered unreliable because the uranium fuel rods' casings — made from the magnesium alloy that gives the design its name — can crack, destabilizing the reactor. And the North Korean device, says Alvarez, is a replica of the Calder Hall reactor down to the last vacuum tube.

The US-designed reprocessing plant is similarly unreliable. Unlike modern versions, the plant requires chemicals to be mixed manually, increasing the chances of spills and leaks. And Alvarez describes the pool where spent fuel is kept as a "radioactive soup". He says that the protective cladding surrounding the spent fuel rods has been compromised, and that the pool water is dangerously radioactive. Uranium in the rods could potentially react with air and ignite.

Despite this apparent decrepitude, the US Central Intelligence Agency estimates that Yongbyon produced 8–16 kilograms of plutonium before it was shut down — enough for at least two weapons. More could be added if the reprocessing facility is reactivated, although Alvarez points out that its materials have been stagnating since 1994. "Think about having your car sitting for that long," he says. "It isn't likely that they can just turn the key and everything will work."

If North Korea has used its plutonium to build a weapon, it is unclear what it intends

to do with it. Retired physicist Richard Garwin, who worked on US nuclear-weapons programmes and was part of a 1998 US commission that assessed the missile capabilities of several countries, thinks that the country may plan to sell nuclear weapons, as it already exports ballistic missiles. Others argue that the nuclear programme is just a tool for negotiating with the United States. "There are only a few people in North Korea who know why they're doing this," concludes Mitchell Reiss, dean of international affairs at the College of William and Mary in Williamsburg, Virginia, who headed the body charged with implementing aspects of the 1994 agreement.

Whatever the reason, Garwin and others agree that Japan and South Korea could feel pressured to develop their own nuclear weapons if North Korea does so. Both countries have the resources to build them—Japan, for example, has up to eight tonnes of plutonium within its borders. Japanese officials say that the fuel is to be used in nuclear power plants—but it could also be used to make a weapon. And although South Korea has signed a deal with the United States saying that it will not produce weapons-grade nuclear material, it does have the resources to develop such weapons should it choose to do so.

Scientific stance

If South Korea and Japan decide to initiate nuclear-weapons programmes, how will their physicists and engineers respond? The attitude of these groups is clearly important. On a practical level, the development of nuclear weapons is impossible without skilled scientists. Garwin points out that in the case of Pakistan and India's nuclear-weapons programmes, which were established in the 1970s, it was the scientists who pushed the technology, while the military was initially uninterested.

On the surface, Japanese and South Korean scientists seem to be opposed to nuclear weapons. The 8,000-member Atomic Energy Society of Japan has pledged not to develop them, mirroring the country's constitutional principle never to develop, hold or use nuclear weapons. "The peaceful use of nuclear energy



A plume of smoke (circled, inset) rising from the Yongbyon nuclear reactor supports suspicions that the complex (main picture) is active again.

has been etched into the Japanese DNA over the past half-century," says Tetsuo Sawada, a nuclear engineer at the Tokyo Institute of Technology. Jungmin Kang, a nuclear engineer who is now a Seoul-based associate of the Nautilus Institute, a security and sustainability think-tank in Berkeley, California, says that South Korean scientists are similarly resistant to nuclear weapons.

But a closer inspection reveals that these attitudes may result from the reluctance of scientists in the region to become involved in politics, rather than a fundamental objection to nuclear weapons. Over the past few years, Tatsujiro Suzuki, an expert in atomic-energy policy at Keio University in Tokyo, has been trying to get nuclear scientists to sign a 'peace pledge', in which they vow never to work on nuclear weapons. By this January, just 110 Japanese researchers had signed.

Those who declined to sign won't say publicly why they did so. A reluctance to declare personal politics may be part of the reason. "Many say that taking an individual stand is

not part of Japanese culture," says Suzuki. But others may be refusing the pledge because they want to leave the door to nuclear weapons open. "Some good friends said they wouldn't sign because they thought they might be asked to work on nuclear weapons in the future," says Suzuki. And when he polled his students, 30% said that their country should eventually develop nuclear weapons. It seems that resistance to nuclear weapons among Japanese scientists may not run very deep.

Suzuki has also had problems in persuading South Koreans to sign his pledge. Kang says that there is little chance of South Korea developing nuclear weapons in response to activities in the North. But if Japan reveals nuclear ambitions, things could change. "I am sure that South Korea will try to develop nuclear weapons if Japan does so," he says. This attitude is reflected in the words of those South Koreans who refused to sign Suzuki's pledge, many of whom cited Japan's plutonium stocks as a reason for not ruling out nuclear weapons.

Such evidence is only anecdotal, but it suggests that if the political leaders of Japan and South Korea want to develop nuclear weapons, they will find scientists willing to help them do so. While North Korea keeps its intentions shrouded in a diplomatic fog, the likelihood of this occurring is unclear. But if the country does push ahead with its plans, those who seek to prevent nuclear proliferation may have to look beyond the scientific communities of neighbouring nations for support.

David Cyranoski is *Nature's* Asian-Pacific correspondent. Geoff Brumfiel is *Nature's* Washington physical sciences correspondent.

Suzuki's peace pledge

► www.peacepledge.gr.jp



North Korea's nuclear reactors use 1950s technology (left); its stores of spent fuel (above) have been called "radioactive soup".

In search of a miracle

Paralysed patients are looking to scientists working on spinal-cord regeneration to help them walk again. Is this pressure causing too much faith to be placed in preliminary, inconclusive results? Helen Pearson investigates.

It was a miracle of almost biblical proportions. In 1998, scientists in Israel revealed that rats whose spinal cords had been severed had walked again after an injection of healing immune cells called macrophages¹. Their hopes buoyed by enthusiastic media coverage, paralysed patients began to dream of taking their own tentative steps.

Five years on, the status of those dreams remains unclear. Proneuron Biotechnologies, a Los Angeles-based company founded on the back of the research of lead investigator Michal Schwartz at the Weizmann Institute of Science in Rehovot, is expected soon to release the results of an initial trial of the procedure on eight patients with spinal injuries. Media reports have indicated that some patients have recovered some feeling and movement, but many researchers do not expect a repeat of the rodent miracle. Indeed, they claim that at least one group has since tried, and failed, to reproduce Schwartz's original animal results.

Schwartz's study is not alone in this regard. Over the past few years, scientists working on spinal-cord repair have revealed encouraging results on several occasions, only to find that other groups have struggled to recreate the same outcome. Three papers published in *Neuron*²⁻⁴ last month underline the point, reporting contradictory findings in parallel studies of 'knockout' mice lacking proteins that are believed to be among the main inhibitors of nerve growth in the spinal cord. "Reproducibility has been a major problem in spinal-cord injury," says Oswald Steward, director of the Reeve-Irvine Research Center at the University of California, Irvine.

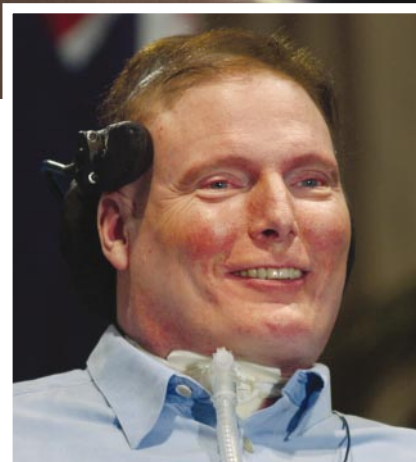


Inspiration: patients with spinal injuries have been given hope by the impetus for research provided by the paralysed actor Christopher Reeve (inset).

Some think that the problems are simply technical: repairing a rat's string-thin spinal cord is a complicated experiment to copy. Others argue that, in some cases, eager scientists have put an over-optimistic spin on their results. These accusations are hotly disputed, but some researchers privately fear that the expectation of high-profile patient-advocacy groups may be inadvertently creating an atmosphere in which problems are likely to occur. Certainly, some researchers admit to feeling under pressure to perform. "If patients phone you three or four times a year and you can't tell them anything new, that's a pressure," says Isabel Klusman of the University of Zurich in Switzerland.

Delicate balance

Exploring these issues is difficult, as no one wants to be perceived as criticizing patient groups. In particular, researchers say that they owe an immense debt of gratitude to the Hollywood star Christopher Reeve, who was paralysed in a horse-riding accident in 1995. His public determination to recover — and his inspirational support of the science that might help him to — has drawn millions of dollars into research, both directly through the Christopher Reeve Paralysis Foundation

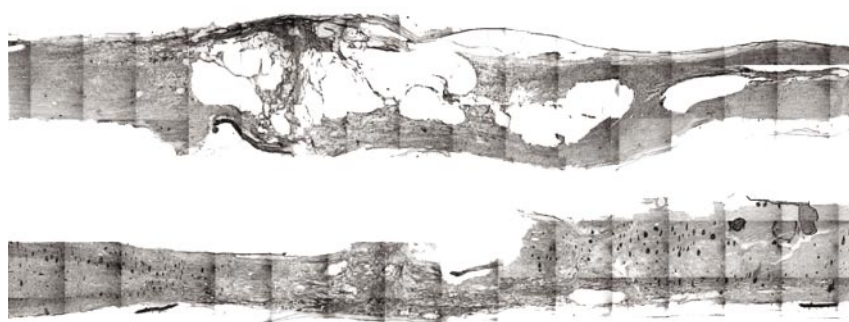


SPL; T. WIMBORNE/REUTERS (INSET)

and indirectly through government budgets.

This impetus intensified interest in a finding reported in 1980, which provided the first glimmer of hope that a damaged spinal cord might be repaired. Albert Aguayo, now at McGill University in Montreal, Canada, cut rats' spinal cords and showed that they were able to sprout into pieces of tissue grafted from the animals' sciatic nerves⁵. This seeded the idea that the lack of regrowth following a spinal injury is partly caused by inhibitory molecules in the spinal cord itself.

Although Reeve's advocacy has undoubtedly brought more funding for research, it has also meant that promising scientific discoveries have been more likely to become



Road to recovery? A damaged rat spinal cord treated with immune cells (bottom) looks to be in better condition almost half a year after injury than cord that was not treated with the cells (top).

front-page news. One study from 1996, for example, received massive coverage. Henrich Cheng at the Karolinska Institute in Stockholm, Sweden, showed that paralysed rats were able to move their legs again after he coaxed spinal nerves to grow across their severed spinal cords. To achieve this, he built 18 tiny bridges of nerves taken from between the ribs and added a growth-promoting protein⁶.

For more than six years, no one published a duplication of the results. Eventually, researchers led by Vernon Lin at the Long Beach Veterans Affairs Medical Center in California managed to reproduce Cheng's main findings⁷. Lin regards human trials as premature, but Cheng has moved ahead, operating on 40 spinal-injured patients in Taiwan. He expects to publish his results later this year.

Technical trauma

Neuroscientists agree that, in large part, the problems with reproducing studies lie in the technical difficulties. "We're trying to do one of the most difficult things in science," says spinal-cord researcher Wise Young of Rutgers University in Piscataway, New Jersey. Cheng says that he took a year to learn how to build the delicate bridges across a severed spinal cord. "We gave up after six months," says Young.

Another problem is that scientists do not always compare like with like. Each group tends to use its own model of injury—including those that crush the spinal cord, or sever it partially or completely. Crushing more closely mimics a typical human injury, but cannot be compared with severing the cord.

Researchers have also struggled to accurately measure and compare improvements in animals' movement after treatment. In another widely debated study, Almudena Ramón-Cueto, now at the Spanish national research council's Institute of Biomedicine in Valencia, reported that cells taken from an adult rat's olfactory bulb, the first staging post in the neural perception of smell, and transplanted into a severed spinal cord, helped paraplegic animals to walk again⁸. Most experts agree that these 'olfactory ensheathing cells', which normally help nerve cells grow projections from the nose to the brain,

can help nerves to grow across a site of injury. But subsequent studies have been difficult to compare with those by Ramón-Cueto, partly because the team assessed the animals' recovery by judging their performance in climbing on a wire ramp—a test not used in other labs.

Although technical difficulties may be part of the problem, some experts argue that the field is also plagued by premature enthusiasm. If, after treatment, only one or two nerves grow across a severed cord, this could be seen as a disappointing result.

Yet, given the normally barren appearance of such a wound, some researchers might describe the same result as a stunning success. "They're seeing what they want to see," claims Fred Gage, a neuroscientist at the Salk Institute for Biological Studies in La Jolla, California.

Researchers whose results have proved difficult to repeat reject any suggestion that they may have been guilty of excessive enthusiasm. "There was no over-interpretation," says Schwartz, "I'm standing behind every word I said." Nir Nimrodi, chief executive of Proneuron, says that the company has replicated and expanded on Schwartz's animal studies, but has not yet published the results, for reasons of intellectual property. Cheng, meanwhile, says that the data from his experiments and the surgical experience of his team gave him the confidence to transfer the technique from the rat model to patients.

Susan Howley, director of research at the Christopher Reeve Paralysis Foundation, based in Springfield, New Jersey, argues that good scientists should not be unduly influenced by patients' desire for good news. Ultimately, she stresses, patients are the losers if preliminary results are overplayed. "They

are on a yo-yo," Howley says—buoyed up by one positive result, only to be crushed by the failure of subsequent studies to repeat it.

How can this roller-coaster experience be avoided? One positive sign is an initiative established last year by the US National Institute of Neurological Disorders and Stroke in Bethesda, Maryland. The \$8-million, five-year Facilities of Research Excellence in Spinal Cord Injury programme will devote a proportion of its budget to determining which animal studies are worth pursuing into the clinic. Independent investigators will be encouraged to work with the labs that first report promising findings to ensure that techniques are comparable. "We need to know what is working and under what conditions," says programme director Arlene Chiu. The initiative will also help to develop more quantitative and objective scales to measure animal recovery.

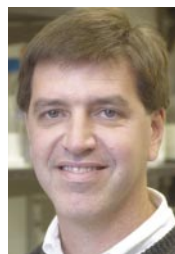
Other experts take solace from the authors of the three papers published last month in *Neuron*, all of which focused on proteins made by a gene called *nogo*. One of the proteins, called Nogo-A, is present in the myelin sheaths that wrap up nerves in the spinal cord, and is thought to help prevent nerves from growing back into an injured site. The three groups, led by Stephen Strittmatter of Yale University in New Haven, Connecticut, Martin Schwab of the University of Zurich and Marc Tessier-Lavigne at Stanford University in California, genetically engineered mice to lack Nogo-A. Strittmatter's group found massive regeneration and improved gait after a spinal-cord cut², Schwab found some regeneration³, and Tessier-Lavigne found none⁴. The reason for the varied results is unclear, but one possibility is that minor differences in the way that the genes were inactivated affected other proteins that block or boost nerve growth.

Rather than racing each other to publish their results separately, allowing patients' hopes to be raised and then dashed, the researchers agreed to work in consultation. Tessier-Lavigne and Strittmatter realized they were doing similar experiments when they met at a 2001 conference, and later approached Schwab. Tessier-Lavigne even sent some animals and a postdoctoral researcher to Strittmatter's lab to try to ensure that they were making the same type of injury.

Mary Bunge, of the Miami Project to Cure Paralysis at the University of Miami School of Medicine in Florida, argues that similar collaborative approaches must be more widely adopted in order to solve the problems of reproducibility plaguing the field. "It's very important that we all get together and talk about it," she says. ■

Helen Pearson works in Nature's news syndication team.

1. Rapalino, O. et al. *Nature Med.* **4**, 814–821 (1998).
2. Kim, J.-E. et al. *Neuron* **38**, 187–199 (2003).
3. Simonen, M. et al. *Neuron* **38**, 201–211 (2003).
4. Zheng, B. et al. *Neuron* **38**, 213–224 (2003).
5. Richardson, P. M. et al. *Nature* **284**, 264–265 (1980).
6. Cheng, H., Cao, Y. & Olson, L. *Science* **273**, 510–513 (1996).
7. Lee, Y. S. et al. *Neurotrauma* **19**, 1203–1216 (2002).
8. Ramón-Cueto, A. et al. *Neuron* **25**, 425–435 (2000).



Marc Tessier-Lavigne (top) and Stephen Strittmatter tried to reconcile conflicting results.

First past the post

From the moment the mysterious illness known as SARS was declared a global threat to health, virologists were racing to develop a diagnostic test. Alison Abbott visits the tiny German lab that got there first.

Christian Drosten is exhausted, but last week he was putting on a brave face for the television crews trying to squeeze into his poky lab. Drosten's fatigue, and his sudden celebrity, both stem from the fact that his team developed the first diagnostic test for severe acute respiratory syndrome (SARS).

Remarkably, Drosten and his colleagues pulled off this feat just 11 days after the World Health Organization (WHO) issued its alert about the disease. And since the test was unveiled on 26 March it has been distributed to more than 150 labs around the world. All in all, it's a considerable achievement for such a small team, given the high-powered virology labs that were engaged in the same quest.

For Drosten, who develops diagnostic tests for viruses and bacteria at the Bernhard Nocht Institute for Tropical Medicine in Hamburg, Germany, the past few weeks have been a whirlwind. The story began in early March, when Drosten and his team — research scientist Stefan Günther and a handful of students — were reading daily Internet postings about a mysterious respiratory illness in Vietnam. "Then, on 15 March, the WHO issued its global SARS alert, and two infected people, a doctor and his wife, landed at Frankfurt airport," Drosten recalls.

Drosten's speciality is polymerase chain reaction (PCR) diagnostics, in which a 'primer' corresponding to a distinctive sequence from a known virus or bacterium is used to amplify the pathogen's genetic material, if it is present in a sample. Initial tests in Frankfurt and elsewhere drew a blank in the search for a viral culprit in sputum taken from the doctor. So when a second sample, taken on 17 March, was sent to the Hamburg

institute to see if a tropical virus might be involved, it fell into Drosten's hands.

Having ruled out the obvious tropical viruses, Drosten began to think about rare viruses that might cause symptoms similar to SARS. He first fixated on the family of paramyxoviruses, whose members include the rare Hendra and Nipah viruses, which can jump from animals to people. "Frankfurt colleagues had looked at the patient's sputum sample under the electron microscope, and the shape, like a squashed sphere, was reminiscent of a typical paramyxovirus," says Drosten.

Think again

Having worked through the night of 18 March on paramyxovirus tests, Drosten got nothing but negative results. He couldn't quite believe it — particularly when Canadian researchers working on samples from patients in Toronto suggested that the causal agent was a metapneumovirus, a type of paramyxovirus. So Drosten sent his sample to Europe's leading paramyxovirus expert, Albert Osterhaus at Erasmus University in Rotterdam, the Netherlands, who confirmed the negative diagnosis.

The remaining option was to try to fish out the elusive virus with a less-specific series of PCR reactions that can amplify the genetic material of a wide range of viruses. This required a pure culture of the virus — the patient's sample would be too full of human genetic material to yield a clean result. Here, Drosten's connections to Frankfurt were crucial. He went to medical school in the city, and many of his former colleagues still work there. And on 20 March, while he was in Frankfurt preparing a presentation with one such colleague, Drosten learned that a culture set up at the university had just begun to yield the virus.

Drosten sped back to Hamburg with a sample on 22 March. His new series of PCR tests took him through a sleepless weekend, during which the machine he was using to sequence the genetic material being amplified failed. But by 25 March, the recalcitrant device had delivered 20 or so sequences.

Two of these matched up with sequences from the coronavirus family. But just as the coronavirus sequence was coming off the machine in Hamburg, researchers at the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, announced that they had identified the SARS agent as a coronavirus. They also had an electron micrograph, which showed the typical crown-like appearance of a coronavirus that had been masked in the Frankfurt image.

Having been pipped by the CDC in implicating a coronavirus, Drosten wasted no time in creating primers that would allow other labs to test for its presence. The next day, he described them on the institute's website and offered to provide them for free. He also made available synthetic sequences that mimic the behaviour of the virus in the PCR test. These 'positive controls' are important, as negative PCR results can often mean simply that the procedure has failed. Although scientists in Hong Kong and at the CDC have since developed similar tests, they have not distributed them as widely, nor provided positive controls.

After such a hectic few weeks, Drosten is looking forward to life returning to normal. His colleagues, meanwhile, are basking in the reflected glory. "But we think he needs a long rest and a good feeding up," says one. ■

Alison Abbott is *Nature's* senior European correspondent.

♦ www.bni-hamburg.de



World-beater: Christian Drosten burned the midnight oil to be the first to produce a rapid test for SARS.

How technology can reduce our impact on the Earth

Prudent use of innovations could avoid sacrificing the present for the future, or vice versa.

Sir — William E. Rees, in his Concepts essay “A blot on the land” (*Nature* **421**, 898; 2003), uses the ecological-footprint concept to argue that the ‘carrying capacity’ of the Earth has been exceeded because of technological and economic growth, and to counter some economists’ claims that the carrying capacity can increase indefinitely. The critical point, unrecognized by either side, is not whether the carrying capacity can increase indefinitely but whether it can increase rapidly enough to accommodate the environmental and economic expectations of a world that grows wealthier as its population growth rate slows dramatically.

Paradoxically, both technology and economic development provide the means to solve the very problems they create. Without technological development in the first instance, the human population would be smaller, because higher birth rates would have been offset by higher mortality rates. Dispensing with present technology now would undoubtedly be catastrophic in human terms — people would be hungrier, unhealthier and shorter-lived, without the world necessarily becoming ecologically more stable.

Similarly, foregoing economic development, which helps to generate wealth, would also be calamitous (see I. M. Goklany, *Case Western Law Review*; in the press). Only wealthy countries can afford the scientific infrastructure to research, develop and put into use clean technologies that increase the Earth’s carrying capacity.

For all of these reasons, the richest countries, not surprisingly, are also the most technologically advanced. They have the highest crop yields per hectare, which is inversely related to the demand for land, a primary element in the ecological footprint. Inefficient agriculture creates pressures for new agricultural land at the expense of virgin forest or marginal lands in countries with growing populations. If agricultural-technology development had been frozen in 1961, we estimate, using data from the Food and Agriculture Organisation (see FAOSTAT 2003: apps.fao.org), that cropland would have had to increase from its present 11% to some 25% of the planetary surface to produce the same amount of food now.

Accepting Rees’s estimate that we currently exceed the Earth’s carrying capacity by one-fifth, without technological development we would now exceed it by one-third. Virtually no natural forest would now remain and the rest of nature

would be even more embattled. Yes, we recognize that current agricultural technology, with its reliance on pesticides and fertilizers, created many new problems even as it solved old ones, but that is exactly why we favour technological change. New technologies need not be perfect, but they should improve on current versions. That is why we support prudent use of agricultural biotechnology — another imperfect technology, but vastly superior to conventional technologies. The trick is not to sacrifice the present for the future, or vice versa.

Without technological change and

economic development, there can be no solution to the predicament of meeting human needs while containing human impact on the planet. Although neither technological change nor economic development is a panacea, they make a solution more likely.

Indur M. Goklany*, **Anthony J. Trewavas†**

**US Department of the Interior, 1849 C Street NW, Washington, DC 20240, USA (The views expressed here are not necessarily those of any branch of the US Government)*

†Institute of Cell and Molecular Biology, King’s Buildings, University of Edinburgh, Edinburgh EH9 3JH, Scotland

Unbalanced view of a dynamic world

Sir — Sandra Knapp, in her Concepts essay about conservation, “Dynamic diversity” (*Nature* **422**, 475), correctly and succinctly presents an analysis of a major conceptual problem facing conservationists: that preserving nature is not about stasis, but about maintaining the ever-evolving variety of life of Earth. Since the middle of the nineteenth century there has been abundant evidence that the Earth has been, and will continue to be, a complex, ever-changing series of habitats and associated species, of which we experience only a snapshot in our lifetimes. Our conundrum is that while we know this, we seem unable to take the required long perspective implied by Knapp’s plea for a dynamic view of the natural world.

Why do we cling to this static view of nature, which is deeply embedded in popular understanding of the natural world as well as in the thinking of politicians and policy-makers interested in conservation and sustainability? This misunderstanding of the natural world can be summed up in a phrase that we constantly hear from the media, environmental groups, politicians and in casual conversation: “the balance of nature”. The idea that nature is, or should be, “in balance” is deeply ingrained. But there is not, and never has been, a balance of nature.

Balance is a seductive concept as it suggests that opposites can be equated and, in the final analysis, traded. But it is not possible to equate humans and elephants, woodlands and meadows, tonnes of carbon emissions with hectares of new forest, and so on, unless it is by reference to an arbitrary, external concept. In the marketplace, this arbitrary concept is money. Economists are now attempting

to balance natural processes, habitats and species using the abstraction of their monetary value, on the assumption that cash can be the final arbiter of the difficult value judgements necessary for custody of the planet and its natural processes.

It is now surely time to abandon this way of thinking. We now need to embrace — in science, popular culture and politics — the phrase “the dynamic diversity of nature”. It may not be as catchy as “the balance of nature”, but wide exposure should eventually see it embedded in our lexicon of aphorisms.

Graham Martin

School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

Lack of trained security staff delays US visas

Sir — Your News story “Researchers rage at tightened restrictions on US immigration” (*Nature* **422**, 457–458; 2003) highlights a growing impediment to international science and technology collaborations. For scientists, students and would-be terrorists wanting to enter the United States, the initial point of contact with immigration policies and practices is the US consular-affairs officer in their own country. This is usually a junior Foreign Service officer — typically new to the country and the local language, lacking in scientific or technical training, and demoralized by long lines of anxious applicants and by highly regulated routine work assignments. Consular Affairs is one of five job categories or ‘cones’ within the State Department. It is the least attractive cone among career Foreign Service officers. Its workload is expected to rise to 12 million non-immigration visa applications by 2005

(according to a report in the *Foreign Service Journal*, March 2001), making visa and passport services an even less attractive profession.

Last August, consular officers received updated State Department instructions on how to apply the Technology Alert List and the list of 'state sponsors of terrorism' to the visa-screening process. Ideally, consular staff should be augmented by intelligence and law-enforcement personnel trained to recognize suspect applicants, and others with scientific or technical backgrounds. Instead, the current officers have been instructed to post these two 'cheat sheet' lists at the "interview windows where the staff can become familiar with the contents". (The department's updated instruction cable is available online at <http://travel.state.gov/state147566.html>.)

The visa staffing, instructions and evolving interagency review process seem guaranteed to further slow the entry of foreign scientists, students and visitors into the United States. For now, official Washington appears to be content with that frustrating backlog, as suggested by congressional comments such as: "Our security is more important than your convenience." That attitude is likely to endure until the airlines, tourism and other global industries begin to make the same complaints that the scientific community is making now.

Edward McSweeney

*US State Department Diplomacy Fellow
(1986–1987), 1692 Barrister Court,
Crofton, Maryland 21114, USA*

Designer scientific literature

Sir — Your News report¹ "Axeing of website article sparks row at Max Planck", describing the removal of several hundred web pages discussing a concept called 'intelligent design' (ID), is welcome.

In Germany, efforts to undermine evolution education — mostly in the form of ID, which rejects the theory of natural selection — have evolved into a successful campaign, including a standard textbook in its fifth edition, several journals and two professional video films in which proponents of ID such as the microbiologist Siegfried Scherer and the geneticist Wolf-Ekkehard Lönnig give interviews in the laboratories of their government-sponsored departments. The ID strategy is not to identify the 'designer' as God in the Bible or for adherents to call themselves creationists; they have coined the term 'theists' to describe themselves (see ref. 2 for a discussion).

Last year, ID-creationism took a step

towards scientific respectability when Lönnig and Heinz Saedler published an review³ entitled "Chromosome rearrangements and transposable elements". In this article they summarize arguments against Darwin's concept of gradual evolution with reference to the prominent German anti-Darwinists Otto Heinrich Schindewolf (1896–1971) and Richard Goldschmidt (1878–1958).

Lönnig and Saedler discuss the possibility of "a partly predetermined generation of biodiversity and new species", which they characterize as a "nonselection-driven and autonomous" process. Popular books by ID proponents Michael Behe and William Dembski are cited as credible sources. (For critical reviews of these books, see refs 4 and 5.) Lönnig and Saedler refer to a "wide range of opinions" and cite evolutionists such as Michael J. Benton, Stephen Jay Gould and John Maynard Smith as well as ID-creationists such as Behe and Dembski, and Lönnig's now-removed web pages. On the basis of these references and polemical comments, the authors state that we should welcome all ideas and hypotheses on the origin of life, "wherever they may lead".

In a German video film called *Is The Bible Right? There is No Evidence for the Theory of Evolution*, Lönnig argues that an intelligent force, endowed with consciousness and spirit, has been at work in the creation of all complex forms of life. This viewpoint is now implicitly proposed as a hypothesis in the scientific literature.³

Four years ago, this journal published two excellent editorials^{6,7} entitled "The difference between science and dogma" and "Combating the exploiters of creationism". I think that the time is ripe to continue this series.

U. Kutschera

*Department of Biology, University of Kassel,
Heinrich-Plett-Strasse 40, 34109 Kassel, Germany*

1. Abbott, A. *Nature* **422**, 460 (2003).
2. Paley, B. A. *Evolution* **56**, 1718–1720 (2002).
3. Lönnig, W. E. & Saedler, H. *Annu. Rev. Gen.* **36**, 389–410 (2002).
4. Coyne, J. A. *Nature* **383**, 227–228 (1996).
5. Charlesworth, B. *Nature* **418**, 129 (2002).
6. *Nature* **400**, 697 (1999).
7. *Nature* **402**, 843 (1999).

Peer review: recognition via year-end statements

Sir — Although I agree with T. Clausen and O. B. Nielsen in Correspondence ("Reviewing should be shown in publication list", *Nature* **421**, 689; 2003) that peer review is a very significant factor in the quality of the scientific literature, their suggestion that peer-review activities should be shown in scientific CVs has practical problems, as they themselves note.

An alternative solution might be for journals to send a letter to their reviewers each year stating how many manuscripts they have reviewed, with some associated measure of quality. This verifiable information should become one criterion for assessment exercises, and would also improve and maintain the general standard of the peer-review system.

The efforts of reviewers should not be underestimated. Even a short and exceptionally well-written manuscript takes at least three or four hours to review properly: more commonly this task takes a day or more. Most manuscripts are revised by their authors and reviewed again by the original reviewers. Two or three reviewers are involved in every manuscript, and about two-thirds of submitted manuscripts are rejected. Hence, on average, each published article has received about 10–15 days of reviewing activity.

Good reviewers may receive one or two manuscripts a month from each journal that knows of them. It follows that these scientists are spending a large percentage of their time on reviewing manuscripts that could otherwise be spent on research, writing and so on (although of course they themselves benefit from peer-review when they submit their own papers).

Editors of journals complain that it is becoming more difficult to attract good reviewers because university researchers increasingly need to earn 'scientific credits'. What is needed, therefore, is a change in attitude from university managers, boards, agencies and others who decide about grants, tenure, promotion and so on.

It has been (and will be) mentioned many times that the current system in which quantity is taken as a measure for academic achievement should be replaced by one that gives credit to quality (*Nature* **422**, 259–261; 2003). Reviewers are chosen because of their quality: their standing in their own discipline and their ability to think critically. An endorsement by a journal could be one way to acknowledge this ability in real terms, by incorporating peer-review activities into the professional career structure. The more prestigious the journal to researchers in the field, the more weight could be given to peer-review activities for that journal by assessment committees.

If reviewers do not receive public credit, the better scientists will eventually no longer be prepared to do this work, which would then devolve to less good reviewers, and the standard of scientific publications would fall.

A. J. (Tom) van Loon

*President, European Association of Science Editors,
Valle del Portet 17 bajo, 03726 Benitachell,
Alicante, Spain*

Going back to our roots

What can genetic techniques tell us about human origins and diversity?

The Journey of Man: A Genetic Odyssey

by Spencer Wells

Allen Lane: 2002. 256 pp. £20

Princeton University Press: 2003. \$29.95

Reflections on our Past: How Human History is Revealed in Our Genes

by John H. Relethford

Westview Press: 2003. 320 pp.

US\$26, Can\$40, £19.99

Rebecca Cann

How can a non-specialist, eager to keep up with the deluge of human genome diversity studies but not attuned to the details and limitations of each particular paper, let alone the context in which they become important, hope to cope? These two books provide a solution. Their intended audiences are rather different, I suspect, but reading a sample from both offers a more balanced understanding of the current state of knowledge than the focus assumed by each author individually.

Spencer Wells chronicles the history of genetic population studies, starting with Darwin's puzzlement over the diversity of humanity he saw first-hand from the deck of the *Beagle*, and ending with the various attempts to classify human variation on the basis of different political and social agendas. This theme is also introduced by John Relethford, who reveals the arbitrary nature of classification schemes imposed by the human mind when conditioned to think about pattern recognition, by using as illustration his failure to answer an IQ-test question correctly.

Both books start with the big picture but rapidly diverge from it. They eventually light on some of the same topics, including Indo-Europeans, Basques, Polynesians and Native Americans, but with different mindsets. Wells assures us that the answers to old questions are at hand, whereas Relethford always cautions prudence. These different approaches to the same topics reminded me of when I was 16 and took my father's new sports car for a spin down a California freeway with my terrified oldest sister next to me in the front seat and my little brother hanging out in the back, yelling and mooning his friends.

Wells offers a quick summary of Richard Lewontin's revelation from 1972 that the human racial classifications commonly used at the time were not founded in modern evolutionary biology. A demonstration of just how creaky our understanding of human diversity, and by extension all biological anthropology, really was first appeared with studies of mitochondrial DNA from living humans. Wells tells the story of how we got



Looking back: geneticists are studying diversity to try to piece together our evolutionary history.

into this mess, and how Luigi Luca Cavalli-Sforza laboured for years to bring order to the field, and honours the work of his mentors (Lewontin was Wells' PhD supervisor and he was a postdoc with Cavalli-Sforza).

But Wells ignores the contributions of their competitors, including Masatoshi Nei, Morris Goodman, Bob Kirk, Naoyuki Takahata, Newton Morton and Robert Sokal, in his haste to cut to the real story he wants to tell. This is the 'inside scoop' on human history provided by the Y chromosome. It reveals the uncomplicated unilineal genetic patrimony of our species, and fits the archaeological record like a dream, except for one tiny detail: modern humans carrying the ancestral M168 Australasian Y haplotype had already expanded into Australia 10,000 years before they swept out of Africa. Never mind, underwater archaeology will eventually clear up the discrepancy.

Wells excels when he explains genetic drift, molecular clocks, climate, the geographic barriers that limited genetic exchange between local groups, and the human Y-chromosome genealogy. In contrast, his assumption of neutrality for Y markers, which would complicate the estimates of when different Y-chromosome groups split from common ancestors, is rather glib. Sexual selection is never specifically mentioned, nor is female mate choice, however cryptic.

Testosterone rules the day: his target audience wants a good 3-hour travel narrative, and he intends to deliver. You won't find the Yagnob people on the Silk Road now, they're

driving taxis in Dushanbe, Tajikistan's capital. This is the MTV version of modern human diversity and genetics. Wells has an insider's knowledge of the science and its excitement, along with a conviction that the African-replacement hypothesis has been proved. But his writing lacks the thoughtfulness of the essays in Steve Olson's *Mapping Human History* (Bloomsbury, 2002), for example, which link ethnicity, gender politics, commercialization of the genome, globalization and gene scans of indigenous people.

Relethford takes a different approach, striving behind the scenes to ensure that readers don't automatically dismiss biological anthropology grounded in morphometrics as data-poor and hypothesis-empty, ready to be picked off by computers full of genotypes. His pleading for the continued examination of the multi-regional hypothesis, in its latest metamorphosis, should be required reading for law students. What constitutes proof in science? People who assume that Y-chromosome DNA tests prove that Thomas Jefferson was the father of Eston Hemings may be surprised.

I enjoyed Relethford's analogy between car makers and haplotypes; the definition of a haplotype (linked genetic markers from the same piece of DNA) is a difficult idea to convey to non-geneticists. How can we be absolutely sure that Neanderthals were replaced, let alone constitute a different species? His explanations of how the chimp genome should inform our judgements about classifying extinct human groups are

clear, as are his discussions of drift and measures of diversity. Some of these ideas will be familiar to readers of his earlier book *Genetics and the Search for Modern Human Origins* (Wiley, 2001).

He does make one glaring error, however, in his discussion of ancient DNA sequences, stating that nuclear DNA degrades too quickly to be reliably extracted from fossils, and that this is why most studies resort to amplification of mitochondrial sequences. He misses the point that it is target template number (2% of the genome bulk is in multi-copy mitochondrial DNA, providing broad comparative possibilities), not quality of DNA, that accounts for the concentration of research effort on maternal gene genealogies.

This book would be an excellent choice for a seminar introducing advanced undergraduates in molecular biology to topics in anthropological genetics, or anthropology graduate students to genetics and historical analyses. Not flashy, far too cautious for my taste, straining to keep the lid on things, he is the exact opposite of Wells. Read them together. ■

Rebecca Cann is in the Department of Genetics and Molecular Biology, University of Hawaii, Manoa, 1960 East-West Road, Honolulu, Hawaii 96822, USA.

Crossing the species barrier

How the Cows Turned Mad

by Maxime Schwartz

transl. Edward Schneider

University of California Press: 2003. 238 pp.

\$24.95, £17.95

Richard E. Race

Until 1996, the diseases of animals known as transmissible spongiform encephalopathies (TSEs), such as scrapie and bovine spongiform encephalopathy (BSE, or 'mad cow disease'), were regarded as agricultural problems with no known medical implications for humans. But then it became apparent that BSE had done what no other animal TSE had done before: transmit to humans. Suddenly, TSE diseases were of great interest to scientists and the public, who wanted to know which other TSE diseases might also pose a risk to people, and how significant the BSE problem would be.

In *How the Cows Turned Mad*, Maxime Schwartz, a molecular biologist at the Pasteur Institute in Paris, develops the history of TSE as a mystery novel. Beginning with clinical descriptions of sheep scrapie three centuries ago, he follows the chronology of TSE evolution through to BSE and its human form, variant Creutzfeldt-Jakob disease (vCJD). Schwartz poses questions, suggests possible



A change in structure turned a harmless prion protein into a fibril that causes scrapie.

answers, and then describes the scientific advances that produced the answers. The book, which was first published in French in 2001, should appeal to a wide range of readers.

Schwartz discusses the basic science behind these discoveries in numerous asides. The general public will find these diversions informative, but those with a reasonable knowledge of TSE diseases might find them distracting. They might also be concerned by the limited bibliography and the oversimplifications that are used to explain many of the controversies that still exist in the TSE field. These oversimplifications include discussions about the nature of the causative agent, the mechanisms of agent replication, the importance of various cell organelles, how infectious aggregates form, and the concept of 'spontaneous' disease. In many cases, seminal research has been omitted and the research of some scientists has been misattributed. Several TSE issues also warrant further comment.

The public's greatest concern regarding TSE diseases is the risk of transmission from animals to people. On this topic, the author is inconsistent, stating on one occasion that TSEs are easily transmitted from one species to another, but elsewhere he suggests that the likelihood is remote. In general, TSE diseases will cross the species barrier if brain homogenates from one species are inoculated directly into the brain of a different species. However, transmission is more difficult or impossible if the material is presented in a natural way, such as by being fed to the second species.

So far, knowledge about which TSE agents are transmissible to which other species has been based on epidemiological observations over the past decades or centuries. But a cell-free analysis that allows the determination of which cross-species transmissions are possible has been developed by my colleague Byron Caughey of the

Rocky Mountain Laboratories in Hamilton, Montana, who has produced results that mimic previous empirical observations very closely. This important work, which is not mentioned in the book, is the basis for recommendations of the World Health Organization and the Centers for Disease Control and Prevention on TSE issues.

The book is also vague about the mechanism by which infectious aggregates of abnormal prion protein are formed. Schwartz mentions two ideas that predominate: one theory suggests that infectious aggregates start off with monomers and dimers of prion protein, whereas the second theory favours a nucleation event. The latter theory is more consistent with the experimental data and is more widely accepted by scientists working on molecular aspects of TSE disease.

A large portion of the discussion involves the idea that, in some situations, normal prion protein can spontaneously convert to abnormal prion protein. Some scientists believe that this phenomenon accounts for cases of sporadic CJD. It seems unnecessary to me to invoke such a mechanism, because it is known that sporadic CJD is infectious and can be transmitted to non-human primates or humans by organ grafts. It is unlikely that we could determine the circumstances of infection, as the incubation periods may be decades and the disease has an incidence of one in a million. This is no reason to believe that an infectious event did not occur.

The author also misrepresents the status of chronic wasting disease (CWD) in wild elk. Although the incidence of CWD in captive or farmed elk can be quite high, it currently affects fewer than 1% of wild elk, even in areas where CWD is endemic. In wild cervids the disease primarily affects mule and white-tailed deer, with variable incidence.

Schwartz wonders why BSE occurred in Britain instead of, or in addition to, other countries with large sheep populations, such as the United States. There are several potential explanations for the lack of BSE in the United States. First, there are far fewer sheep in the United States than in Britain. Second, scrapie has been present in US sheep populations in low numbers only since 1947, compared with more than 200 years in Europe. Finally, the United States has had scrapie-eradication programmes in effect since 1952.

When Carleton Gajdusek won a Nobel prize in 1976 for his discovery of the transmissibility of kuru, a human TSE, the nature of the infective agent was unknown. Another Nobel prize was awarded to Stanley Prusiner 21 years later for his 'protein-only' theory of TSE aetiology. In his discussion of this theory, Schwartz mentions experiments that support the idea, but leaves out data that suggest other possibilities. He points out that even though the critical experiments to prove the theory have not yet been successful, the

idea has nonetheless been widely accepted by the public and even by scientists who are not familiar with TSEs. I would simply re-iterate that the protein-only theory is still just that: a theory. Several prominent TSE scientists believe that the initial conversion of normal prion protein to an abnormal form is not dependent on protein alone. Perhaps a third Nobel prize in the TSE discipline will be awarded to whoever performs the critical experiments that unequivocally prove or disprove the protein-only theory. ■

Richard E. Race is at the National Institute of Health Laboratory for Persistent Viral Diseases, Rocky Mountain Laboratories, Hamilton, Montana 59840, USA.

More on prion diseases

The Pathological Protein: Mad Cow, Chronic Wasting, and Other Deadly Prion Diseases

by Philip Yam

Springer: 2003. £21, \$27.50, £29.95

Copernicus Books: 2003. \$27.50

A collection to remember

Memory From A to Z

by Yadin Dudai

Oxford University Press: 2002. 338 pp. £37.50, \$65

Larry R. Squire

Like many topics in contemporary neuroscience, the study of memory is being approached using a variety of techniques and at many levels of analysis. Work ranges from the molecular and cellular biology of synaptic change in cell culture and brain slices to the analysis of memory in behaving organisms. Mindful of the diverse backgrounds of those who come to the study of memory, neuroscientist Yadin Dudai has written a charming book designed especially for the student. "This book," he writes, "contains terms that I wish my students to know."

What follows is an engaging, informative and sometimes playful collection of essays, each 2–3 pages long, around the topic of memory, arranged in alphabetical order. The topics do move from A to Z (from *a priori* to *Zeitgeist*), but the book is a personal selection of keywords and starting points for discussion, not an encyclopedia. Many letters have multiple entries (there are 20 for C and 13 for P, for example); seven letters (J, K, Q, U, V, X and Y) have no entry at all.

Most of the entries are ones that might be expected in a book about memory and neural plasticity ('synapse', 'glutamate', 'calcium', 'CREB', 'hippocampus', 'long-term potentiation', 'declarative memory', 'forgetting' and 'working memory'), but many of them are

Cinema

Return of the mutants

Science fact is the loose basis for the fantasy in the current spate of movies based on comic-book characters. Peter Parker was bitten by a genetically modified spider last year to become Spider-Man (in the original 1960s' comic the spider was radioactive, but that was a different age).

Director Bryan Singer's latest film, *X-Men 2* (Twentieth Century Fox), is no different, ending with a flowery but accurate description of evolution by punctuated equilibrium. This second adaptation of the popular comic *X-Men* opened worldwide at the beginning of May.

The X-Men are a band of superheroes (*Homo sapiens superior*) who possess a mutated X gene, which has an extraordinarily variable phenotype, allowing some mutants to walk through walls, some to shoot ice from their fingers, and some to perform Moses-like acts of water telekinesis. (The X gene's normal function is not revealed.) The team is led by a powerful telepath, Charles Xavier (played by Patrick Stewart), a benevolent leader who promotes the integration of the mutants into an otherwise hateful society.

Xavier's gang faces two arch-rivals: a human bent on wiping out all of the mutants, and a concentration-camp survivor, Magneto (Ian McKellen), who can control magnetic fields to the point of being able to extract iron from



Cutting-edge science? Remarkable powers of regeneration from a mutated X gene make Wolverine a formidable fighting force.

blood. Magneto fervently believes that mutants are the natural inheritors of the Earth and that mankind should voluntarily go extinct to bring this about.

Themes of biological determinism are touched on in this fun, but perhaps overlong, film. Magneto encourages a young mutant to join him, telling him that his genetic abilities render him "a god amongst insects". The wise Xavier offers a counterpoint later: when his team question why one of their number commits an act of self-sacrifice, he weightily declares that the protagonist, against her nature, "made a choice". ■

Adam Rutherford is the web editor of Nature.

unexpected or even amusing. The entry on 'capacity' provides a thoughtful discussion of what is sometimes called the impossible calculation: how much information-storage capacity does the brain actually have, and does this place any constraint on how much information can be acquired during a lifetime? Entries such as 'artefact', 'controls', 'red herring' and 'Ockham's razor' consider aspects of the scientific method and how science gets done. The entry on 'scoophobia' (fear of being scooped) presents a humorous discussion about issues of competition in science.

The entry on 'palimpsest' is perhaps a stretch, but it forms the basis for a brief account of how new memory representations form in interactions with previously established ones. There are also separate entries for the organisms that have been important to the scientific study of memory: *Aplysia*, *Drosophila*, bird (under 'birdsong'), mouse, rat, monkey and *Homo sapiens sapiens*.

What makes all this work is that the 131 essays that make up the book are scholarly, up-to-date, even-handed and rich in citations

(there are 66 pages of references at the end of the book, along with a full index). Many of the entries begin with an explanation of the classical origins of the terms. Any one of the entries could serve as the basis for classroom discussion and as a starting point for further reading, even for the practising scientist.

A charming feature of the book is that the personal and whimsical list of entries encourages one to try out new ones. For the unused letters, I developed junk DNA, kinase, qualia, quisqualate, ubiquitin, X-rays, *Xenopus*, vestibulo-ocular reflex, virus and the Yerkes Primate Research Center.

In the end, the book describes an associative network of ideas and facts that, for pedagogical purposes, could be extended and recombined indefinitely. This is an engaging format, and it would be interesting to see it tried with other topics. ■

Larry R. Squire is at the Veterans Affairs Healthcare System and the Departments of Psychiatry, Neurosciences and Psychology, University of California, San Diego, California 92161, USA.

Dreams of the past

What was your first experiment as a child?

As a young shepherd boy in *la France profonde* during the Second World War, I had the chance to observe the behaviour of domestic animals — horses, cows, goats and my own dog.

What makes a good scientific mentor?

Someone who believes in you, who is very open-minded, who looks for opportunities at the interfaces of disciplines — and with whom you can hold frank and open discussions.

Whose graduate student would you most like to have been (historical impossibility notwithstanding)?

Charles Darwin, during the voyage of the *Beagle*.

What single scientific paper or talk changed your career path?

The talk that Louis Leakey gave at the Muséum National d'Histoire Naturelle in Paris in 1963, in which he introduced the new species *Homo habilis*.

What book has been most influential in your scientific career?

Two books by Charles Darwin: *On The Origin of Species* (1859) and *The Descent of Man* (1871).

What gives you the most job satisfaction now?

All the young colleagues, PhD students and postdocs in my team: they are the future. I'm particularly proud to train students from Chad and to help build a palaeontology research school in the University of Chad and the Centre National d'Appui à la Recherche in N'Djaména. Of course, it's also very exciting to have found the earliest hominid west of the Rift Valley, but the conditions — in the Djurab desert — are extreme, with a lot of sand storms. In that kind of environment, you get to know your own capabilities, and those of the other members of your team, which becomes a large, second family.

What are your major frustrations?

The knowledge that I will never live long enough to achieve all that I want to, and that much of what remains must be spent on administration.

What was the worst/most memorable comment you ever received from a referee?

For the same article, one referee complained about the use of a name I had given in memory of a lost best friend, which he thought "too cute"; and another wrote that it was the most important discovery in the past three-quarters of a century.

What's your favourite conference destination, and why?

Anywhere in the company of children and young people, because you can make them dream with the past, for the future. But, of course, my favourite location is N'Djaména in Chad because of the interaction with all the enthusiastic young Chadians dreaming and thinking about a Chadian cradle of humanity.

What literary character would you employ as a postdoc?

A science-fiction author such as Michael Crichton, because dreaming is always a big part of research.

What book is currently on your bedside table?

Timeline, a science-fiction novel by Michael Crichton.

Where and when would you most like to have lived or worked?

In a university with a research centre that ranked as among the most famous in the world, which could offer me the best conditions ever dreamed of for working; somewhere in *la France profonde*; and right now, when the biological sciences have such a promising future.

You are on a plane behind two students obviously going to the same conference, who start to talk about your work. What do you do?

I would say: "Excuse me, I'm Michel Brunet. Nice to meet you." Then it would be up to them if they wanted to talk further or not.

What's the best piece of advice you've ever received?

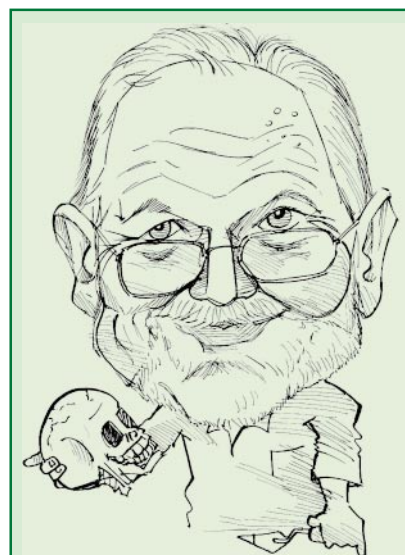
That fraternity is more important than everything else in a world of violence and individualism. Is it not a great message that our origin is African, unique and that we are all brothers and sisters?

The Internet is the bane of scientists' lives because...

...the principle of freedom of expression is not always a guarantee of quality. On the other hand, the Internet gives you the opportunity to communicate in real time with people all over the world.

What overlooked or underrated discovery really changed the science in which you work?

The discovery of the Taung child in South Africa (1924) and its publication, as the first prehuman *Australopithecus africanus* (R. A. Dart *Nature* **115**, 195–199; 1925), was widely dismissed at the time, but history has shown it to have been a landmark.



Michel Brunet

Michel Brunet is at the University of Poitiers, France. A nomadic digger of bones, he is the proud parent of two of our forefathers excavated from the desert of Chad in central Africa, Abel and Toumaï (aged 3.5 million and 7 million years, respectively).

What do you do to relax?

Walk and dream in *la France profonde*.

What would you have become, if not a scientist?

A farmer rearing horses and cattle in the countryside.

What single discovery, invention or innovation would most improve your life?

The use of stents in cardiac surgery kept me alive after my heart attack. For the future, I look forward to a small prosthetic heart with the same efficiency as a biological heart.

What music would you have played at your funeral?

Vivaldi, *The Four Seasons*.

Is there a 'tyranny of reductionism' in how scientists are trained today?

There is not enough about the history of science in our students' academic courses. In addition, even if there is a consensus about transdisciplinary research, young scientists have had to specialize too early in their careers. For instance, it is quite obvious that palaeontological fieldwork requires you to be an excellent geologist as well as an anatomist.

What is the one thing about science you wish the public understood better?

That the 'truth' is always an asymptotic ideal.

Name one extravagance you can now get away with because of your eminence.

Nothing — I don't want to change. ■

Catch a falling star

Jack Drummond

The European Fireball Network has recovered a meteorite after photographing its violent entry into the Earth's atmosphere. This is only the fourth such recovery, and analysis points to a surprising past for this primitive object.

Meteoroids, or rocks from space, are seen travelling through the Earth's atmosphere as fireballs — extremely bright shooting stars or meteors — and any surviving debris is scattered as meteorites. For more than 40 years, networks of cameras have been primed to capture images of the incoming fireballs, and pinpoint the landing sites of their meteorites, with the aim of tracing the precise orbit and likely origin of the meteoroid. But in all that time, meteorites have been recovered from only three detected fireballs. On page 151 of this issue¹, Spurný, Oberst and Heinlein report the fourth instance of meteorite recovery, from the European Fireball Network — the project that also made the first recovery. All four meteorites, with known paths to their final resting place, are unique tracers in time and space of the history of our Solar System — although the latest has also thrown up a conundrum.

In the early years of the twentieth century, Ernst Opik led expeditions to the southwest deserts of North America to make systematic visual observations of meteors. He initially concluded that most meteors were of interstellar origin. In New Mexico, in the late 1940s and early 1950s, Fred Whipple and colleagues made photographic studies using a rotating shutter in front of fast cameras to measure the meteors' velocities. In what is some of the most precise work even by today's standards, these experiments showed that most meteors in fact originated from comets, but that a substantial fraction were in asteroid-like orbits — that is, the furthest points of their orbit from the Sun (their 'aphelia') fell in the asteroid belt, between Mars and Jupiter.

Shortly thereafter, astronomers began to classify asteroids on the basis of the spectrum of light that they reflected². Although some order could be divined from the colour groupings, chemical analysis of meteorites was revealing more about the history of the Solar System than the asteroids could. For example, from the classification of meteorites it became apparent that there are only tens to hundreds of parent bodies for the thousands of meteorite specimens in collections around the world³. Historically, attempts to match the spectra of meteorites to specific parent asteroids have not been easy. But it was realized that if one could pick up a fresh meteorite, look skywards, and know precisely

Figure 1 Shooting star. In April 2002, the Streithelm camera of the European Fireball Network in southern Germany took this time-lapse photograph (top right), capturing the 91-km trail of the meteor as it arrived in the Earth's atmosphere. Three months later a 1.75-kg meteorite (below), part of the fireball's debris, was recovered near the fairytale castle of Neuschwanstein, Bavaria. Spurný *et al.*¹ have analysed the trajectory of the meteor and the chemical make-up of the meteorite. Despite the similarity of its orbit to an earlier recovered meteorite, the Neuschwanstein specimen is chemically very different, suggesting that there is a greater variety of objects in the meteor stream than had been thought.



JOSEF BECK/GETTY IMAGES

where it came from, a big piece of the puzzle might be uncovered. As a result, several networks of cameras, spanning areas of a million square kilometres, were set up in Europe, Canada and the United States. The intention was to photograph fireballs to calculate their orbits and find the meteorites that dropped from them. But the booty from these networks, despite years of sky-watching, has been disappointingly small — just four meteorites in 43 years.

On the evening of 7 April 1959, Zdenek Ceplecha, founder of the European Fireball Network, was watching television when a tremendous fireball lit up the walls of his living-room. He immediately ran to the television set and adjusted its brightness level to produce the same level of illumination of the walls. Calibrating this later, he was able to estimate the fireball's brightness

as being of magnitude -19 , much brighter than the full Moon.

That historic night, his fireball network had photographed and tracked the meteor's path through the atmosphere, and its orbit in the Solar System was calculated. Because the fireball's atmospheric trajectory was known precisely, an estimate could be made of where meteorites might have fallen. A search was mounted and two days later a farmer found a 4.5-kg piece in his field near Příbram, Czechoslovakia. The total mass of meteorite recovered at Příbram amounted to 5.8 kg, from an initial mass of perhaps a tonne (estimated from the fireball's brightness). Příbram is classified as an H5 chondrite — a primitive, very old, undifferentiated stone containing small spheres of silicates (chondrules) reflecting the make-up of the presolar nebula that became the Solar System. Its orbit

had an aphelion in the main asteroid belt and its perihelion was inside the orbit of the Earth, one-fifth of the way to the Sun.

On the night of 3 January 1970, another brilliant fireball lit up the sky over the central United States. As bright as the full Moon (magnitude -12), the meteor was photographed by the Prairie Fireball Network, and its orbit determined. On 10 January, the network team found a 10-kg meteorite lying on top of the snow in the middle of a road near Lost City, Oklahoma. Guided by a local resident who reported seeing glowing red lights and hearing a thud nearby, they found three other pieces, raising the total recovered mass to 17 kg, from an estimated pre-atmospheric mass of 75–200 kg. The Lost City meteorite is also classified as an ordinary H5 chondrite, and its orbit took it from the main belt to just inside the orbit of the Earth.

On 5 February 1977, another fireball of magnitude -12 was reported by an airline pilot flying over Saskatchewan, Canada. Members of the Meteorite Observation and Recovery Project recruited local youths to search for meteorites in the snow, and they found the first specimen near Innisfree, Alberta, on 17 February. Five more pieces were found in April after the snow had melted, and a farmer turned up another three. The total weight of all nine specimens came to 4.6 kg, from an estimated pre-atmospheric mass of 20–40 kg.

And that was it. No other meteorites had been recovered by fireball networks, until now. As Spurný *et al.*¹ report, the fourth meteorite was found on 14 July 2002, 6 km from the famous castle of Neuschwanstein (Fig. 1), after being photographed on 6 April

2002 by the successor to the original European Fireball Network. The fireball was of magnitude -17 and carved a luminous path 91 km long across the sky. Amazingly, it has been found to have an orbit identical to the Příbram meteorite, the first ever recovered (Fig. 3 on page 152). And yet, unlike Příbram (a fairly ordinary chondrite), Neuschwanstein is classified as an 'EL6 enstatite chondrite': it differs significantly from the other three in that it is more reduced (with less FeO in its silicates), it has lower ratios of Mg/Si and (Ca, Al, Ti)/Si, and it has a coarser granularity with little or no evidence of chondrules. Furthermore, Neuschwanstein has a cosmic-ray exposure age—a measure of the length of time that the body has been travelling through space—of 48 million years, compared with 12 million years for Příbram. So it seems that the meteor stream in which these objects were flowing is more heterogeneous than is usually assumed.

Looking skyward, knowing where these rocks came from, we may begin to understand something of the early history of our Solar System. Zdeněk Ceplecha, now retired, must be delighted to see this new success of his fireball network. ■

Jack Drummond is at the Starfire Optical Range, AFRL/DES, 3550 Aberdeen Avenue SE, Kirtland AFB, New Mexico 87117-5776, USA. e-mail: jack.drummond@kirtland.af.mil

1. Spurný, P., Oberst, J. & Heinlein, D. *Nature* **423**, 151–153 (2003).
2. Chapman, C. R., Morrison, D. & Zellner, B. *Icarus* **25**, 104–130 (1975).
3. Burbine, T. H., McCoy, T. J., Meibom, A., Gladman, B. & Keil, K. in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 653–667 (Univ. Arizona Press, Tucson, 2003).

Mathematics

Conjuring with conjectures

Ian Stewart

The Clay Mathematics Institute is offering a million dollars for a solution to the Poincaré conjecture, and Grisha Perelman may have found one. What is the conjecture, and why does it matter?

In 1904, Henri Poincaré was making fundamental advances in topology—the multi-dimensional study of mathematical properties such as 'knotted' or 'connected', which are unchanged by continuous deformations. Buried in his work was an unjustified assumption about three-dimensional spaces. When he noticed it, and failed to find a proof, the assumption became first a question, and then a conjecture. Unsolved almost a century later, the Poincaré conjecture joined other problems (such as the Riemann hypothesis and Yang–Mills theory) on a list of 'Millennium Prize Problems' for whose solutions a substantial cash prize is offered¹. Grisha Perelman of the Steklov Institute of

Mathematics in St Petersburg, Russia, may now have found the answer^{2,3}.

The Poincaré conjecture can be understood by analogy with the case in two dimensions. A two-dimensional space, or surface, is like a bubble made from an infinitely thin film of soap. If the bubble is round, the surface is a sphere, but many other shapes are possible. A torus, for instance, is shaped like an American doughnut; a two-holed torus is like two doughnuts joined together, and so on.

The early topologists proved that any surface is topologically equivalent to a torus with a finite number of holes. A lumpy potato is the same as a sphere, a coffee-cup is the

same as a torus, and a teapot is the same as a two-holed torus⁴. Poincaré sought a comparable understanding of three-dimensional spaces. The simplest is the three-sphere, analogous to a sphere but having three dimensions instead of two. This is almost the same as a solid ball, but only if you pretend that the entire outer surface of the ball is a single point.

The key problem is to decide whether a given surface—a lumpy potato, say—is equivalent to a sphere, or to something more exotic. Poincaré's predecessors had solved that problem in two dimensions, by working out how closed loops in the surface behave when they are moved. Any loop on a sphere can be continuously deformed, or shrunk, to a point. But on a torus, or a more complicated surface, many loops cannot be shrunk (Fig. 1a). Better still, if every loop can be shrunk then the surface must be a sphere. So the 'shrinkability' of loops provides a definitive test for spherical topology.

What about three dimensions? It is easy to prove that every loop in a three-sphere can be shrunk, but Poincaré tacitly assumed the converse: if every loop can be shrunk, then the space is topologically a three-sphere. Only later did it dawn on him that this statement—now called the Poincaré conjecture—was not obvious, and perhaps might not even be true.

Throughout the twentieth century, the conjecture remained as mysterious as ever. Analogous statements in all numbers of dimensions from four upwards had been proved true, but Poincaré's original three-dimensional problem still held out. The Poincaré conjecture had become a serious obstacle to any understanding of three-dimensional topology. Because topology is fundamental to many branches of mathematics, and to several areas of mathematical physics, this missing piece of the mathematical tool kit had become a serious embarrassment. How can we hope to understand the shape of our own Universe, for instance, when we don't even know the range of possibilities? And what about the even harder topological problems arising in quantum field theory (for example, in relation to string theory, where space-time has extra dimensions that curl up in some topological manner)?

Around 1983, William Thurston had devised an entirely new approach to the Poincaré conjecture, by relating it to classical geometry. Starting, again, in two dimensions, Thurston wondered why there are lots of shapes for a potato, but only one round sphere. What makes that shape special? It has constant (positive) curvature: every bit of a round sphere is bent the same amount as any other. Similarly, there is a 'canonical' geometry for a torus, and in this case the curvature is zero. The usual torus does not look flat, but a rectangle with opposite edges

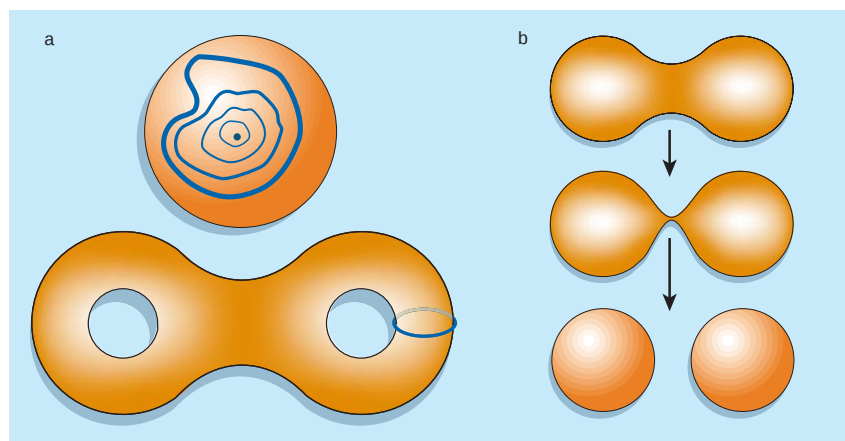


Figure 1 Testing topology. a, On the surface of a sphere, every loop can be deformed to a single point. But on any other surface, such as the two-holed torus shown here, some loops cannot be deformed to a point. Topologically, these are two-dimensional objects, but in three dimensions the 'shrinkability' of loops becomes the basis of Poincaré's conjecture. b, Perelman's solution^{2,3} to the Poincaré conjecture starts with 'Ricci flow': to achieve constant curvature of its surface, an object such as this dumbbell-shaped bubble splits into two round ones.

identified is flat, and topologically that is a torus. Finally, tori with two or more holes can be realized as surfaces of constant negative curvature.

Thurston suggested that something similar should happen in three dimensions. He proved that exactly eight different 'geometries' can occur. It turned out to be too much to expect every three-dimensional space to have a single canonical geometry, but Thurston's 'geometrization conjecture'

asserted the next best thing: every three-dimensional space can be cut up, in a systematic way, so that each piece has precisely one of those eight geometries. This new conjecture was much more ambitious than Poincaré's: it aimed at understanding all three-dimensional spaces, not just the three-sphere. In particular, the Poincaré conjecture was an easy consequence of the geometrization conjecture: the condition on loops meant that only one piece would be

needed, and the associated geometry must be that of the three-sphere.

Perelman's work^{2,3} is a new approach to the geometrization conjecture. If it pans out, it will constitute a huge leap forward in three-dimensional topology and mathematical physics. It is based on the Ricci flow, an idea of Richard Hamilton's⁵: if a soap bubble is deformed away from its normal round shape, surface tension will pull it back to a perfect sphere. The Ricci flow is an analogous way to deform any surface so that its curvature 'tries' to become constant. Along the way, though, it may have to split into separate pieces; for example, a dumbbell-shaped bubble must break up into two round ones (Fig. 1b).

Hamilton defined the Ricci flow in three dimensions, and developed a programme to prove that when a three-dimensional 'bubble' follows the flow, the pieces it breaks into are essentially those predicted by the geometrization conjecture. Perelman's papers do not carry out the entire Hamilton programme, but it looks as though they might establish enough of it to prove the geometrization conjecture. That in turn will prove the Poincaré conjecture.

If everything hangs together, a 99-year search is at an end. If not, Perelman's papers will still shed a huge amount of light on the Ricci flow, which is important in quantum field theory and pure mathematics, and may well pave the way to an eventual proof of the two conjectures. Either way, the spin-off is

Molecular physiology

Tuned for longer life

Sometime early in the sixteenth century, 40-year-old Luigi Cornaro decided to cut his food intake dramatically for health reasons — he then lived on to the age of 102, writing treatises on the merits of his abstemious lifestyle. Five centuries later it is not clear whether severely limiting food consumption can, in general, extend human life, although calorie restriction does indeed extend the lifespan of commonly studied organisms such as yeast, nematode worms and mice. The latest turn of events is described by David A. Sinclair and colleagues elsewhere in this issue (R. M. Anderson *et al.* *Nature* **423**, 181–185; 2003). In studies of the yeast *Saccharomyces cerevisiae*, the researchers have pinpointed an enzyme, Pnc1, that seems to be a key link between environmental stress, metabolic energy and lifespan.

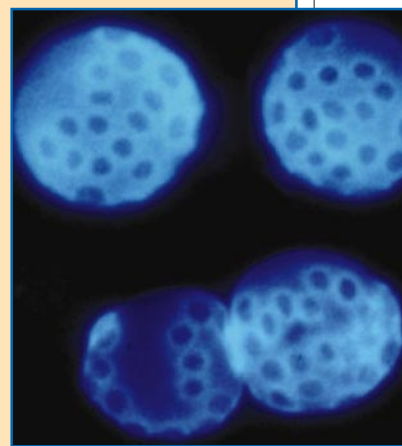
Sinclair and colleagues explore

the control of an intriguing protein named Sir2 (silent information regulator 2), which has been shown to extend the lifespan of various laboratory organisms. Sir2 is a histone deacetylase, found in cell nuclei, where it is thought to turn genes on and off by influencing the packing of chromosomal DNA. Sir2 action requires the ubiquitous energy cofactor NAD⁺, generating nicotinamide in the reaction, and nicotinamide also turns out to be a potent inhibitor of Sir2. But how exactly does metabolic energy affect Sir2 activity and lifespan?

The involvement of nicotinamide led Sinclair and co-workers to Pnc1, an enzyme that breaks nicotinamide down to nicotinic acid. They find that increasing the levels of Pnc1 in yeast leads to a dramatic lengthening of lifespan, as nicotinamide depletion results in Sir2 activation (yeast lifespan is usually measured by the number of

buds, or daughters, produced by a given cell, individual bud scars showing up as darker circles on the blue-stained yeast cells shown here). Various types of stress are known to extend yeast life, including glucose restriction, high temperatures and elevated salt concentrations, and Pnc1 levels are raised in all of these circumstances. This suggests that Pnc1 is a key regulator of Sir2, and is responsible for tuning a yeast's lifespan and reproductive capacity to the quality of its environment.

Whether these ideas apply to multicellular organisms will have to be tested in worms and mice. On the one hand, one might imagine that mechanisms designed to protect single yeast cells from the vicissitudes of their environment would not be needed in more complex organisms, which possess homeostatic systems designed to protect and nourish individual cells.



On the other, lifespan is likely to be controlled by evolutionarily ancient mechanisms. So understanding Sir2 regulation in multicellular organisms, and identifying the genes that are controlled by Sir2 and its relatives, should prove generally rewarding.

Richard Turner

likely to make profound changes to how we think about topology, space-time and quantum field theory.

Ian Stewart is at the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK. e-mail: ins@maths.warwick.ac.uk

1. http://www.claymath.org/Millennium_Prize_Problems
2. Perelman, G. Preprint math.DG/0211159 at <<http://arXiv.org>> (2002).
3. Perelman, G. Preprint math.DG/0303109 at <<http://arXiv.org>> (2003).
4. Stewart, I. *Flatterland* (Macmillan, London, 2001).
5. Hamilton, R. S. *J. Diff. Geom.* **17**, 255–306 (1982).

Prion diseases

Cannibals and garbage piles

Adriano Aguzzi and Mathias Heikenwalder

There's a lot of disagreement among prion scientists, as a recent conference made very clear. Even the revered 'prion hypothesis' came under attack.

After more than 280,000 mad cows and two Nobel prizes for research on transmissible spongiform encephalopathies, we must know all there is to know about these 'prion' diseases. Or do we? This question was asked at a recent symposium*, which looked at topics such as the cell biology of the normal prion protein and of its misshapen, disease-associated form; how the brain becomes damaged; and diagnosis and treatment. It was clear that, in all of these areas and more, there's still a long way to go.

Prion scientists have a reputation for being a contentious bunch — a fact that was amply confirmed on this occasion. Diametrically divergent opinions emerged on central questions such as the physiological function of the normal prion protein (PrP^C) and the role of its aberrant form (PrP^{Sc}) in disease.

Even the prion hypothesis, which nowadays is often regarded as dogma, was challenged. This hypothesis states that PrP^{Sc} is the infectious agent in transmissible spongiform encephalopathies (TSEs), and that it replicates by imparting its misshapen conformation onto PrP^C. In a spirited lecture, however, new Nobel laureate Kurt Wüthrich (ETH, Zurich) pointed out the continued failures to create infectivity *in vitro* by modifying bacterially expressed prion protein — a crucial prediction of the prion hypothesis. Another important experiment involves abolishing the structure (and infectivity) of the disease-associated prion protein with specific salts, and then attempting to restore infectivity by reinstating the original structure. This, too, has so far failed. Wüthrich referred to PrP^{Sc} as simply a build-up of "garbage", and submitted that we must understand the function of the normal prion protein before we can understand prion diseases.

The TSEs are characterized by the death of nerve cells, and another point of controversy concerned the mechanisms by which

this occurs. A strong case was presented that the accumulation of PrP^{Sc} within the cytosol of neurons is to blame (S. Lindquist, Whitehead Inst., Cambridge, Massachusetts)^{1,2}. PrP^C is usually located in the plasma membrane, and travels there by way of a network of internal membranes, the endoplasmic reticulum (ER). Lindquist proposed that a certain proportion of PrP^C never reaches the plasma membrane, but instead re-enters the cytosol by 'retrotranslocation' from the ER — a standard means by which other misfolded proteins are directed to the cell's waste-disposal unit, the proteasome. Lindquist found that a form of the prion protein that was specifically targeted to the cytosol caused rapidly lethal neurodegeneration in mice. This protein did not acquire resistance to protein-digesting enzymes (proteases), which has long been thought to be a key characteristic of PrP^{Sc}. But proteasome inhibition led to the accumulation of a slightly protease-resistant prion protein in cultured cells. Lindquist speculated that the diverse mutations in the prion protein that are associated with familial Creutzfeldt–Jakob disease (CJD) — a human TSE — might all lead to enhanced retrotranslocation, which, upon impaired proteasome function, could trigger disease. The big surprise here is that the cytosol might be the place where PrP-mediated neuronal death begins.

However, this model was contested by D. Harris (Washington Univ., St Louis, Missouri), who found that cytosolic prion protein retains its 'signal peptide' — normally removed after proteins enter the ER — and does not contain the glycosyl phosphatidylinositol 'anchor' needed for attachment to membranes³. This suggests that the protein never entered the ER, and so could not have undergone retrotranslocation. Harris also pointed out that proteasome inhibitors have powerful effects on the levels of prion messenger RNA; these effects might have contributed to previous results.

Even within a single population of inbred animals, prions come in distinct varieties, which — upon transmission to further



100 YEARS AGO

Further particulars of the work and position of the National Antarctic Expedition have been brought by the New Zealand mail...

The chief scientific work accomplished by the expedition is summarised as follows:—

- (1) The discovery of an extensive land at the east extremity of the great ice barrier.
- (2) The discovery that MacMurdo Bay is not a "bay," but a strait, and that Mounts Erebus and Terror form part of a comparatively small island.
- (3) The discovery of good winter quarters in a high latitude — viz. 77° 50' S., 166° 42' E. — with land close by suitable for the erection of the magnetic observatories, &c. The lowest temperature experienced was 92° of frost Fahrenheit.
- (4) An immense amount of scientific work over twelve months in winter quarters, principally physical and biological.
- (5) Numerous and extensive sledge journeys in the spring and summer, covering a good many thousand miles, of which the principal is Captain Scott's journey, upon which a latitude of 82° 17' south was attained, and an immense tract of new land discovered and charted as far as 83° 30' south, with peaks and ranges of mountains as high as 14,000 feet... As the *Discovery* has not returned to Lyttelton, there is little doubt that the expedition has been forced to spend a third winter in the Antarctic.

From *Nature* 7 May 1903.

50 YEARS AGO

In June 1952 the crew of a fisheries patrol boat cruising in Miramichi Bay, New Brunswick, observed what they took for a sand-bar in an unexpected situation. On closer inspection they found that there was no such bar, but that the surface of the sea over a wide area was in a state of violent turmoil ("a sort of boiling") due to the presence of enormous numbers of polychaetes. The area covered was some 150 yards in radius and was broken up into several minor patches... The polychaetes involved were the heteronereid forms (both sexes) of *Nereis succinea* (Leuckart). This is the first record of that species from eastern Canadian waters, though it is known from the eastern coast of the United States, usually under the name *Nereis limbata* Ehlers... We know of no other case of the swarming of a nereid with such activity and in such concentrated masses as occurred in this instance.

From *Nature* 9 May 1953.

*Keystone Symposium on Molecular Aspects of Transmissible Spongiform Encephalopathies (Prion Diseases), Breckenridge, Colorado, 2–6 April 2003.

animals — produce characteristic, heritable incubation times and patterns of brain damage. This phenomenon of prion 'strains' continues to produce surprises. When hamster prions were inoculated into mice, the animals lived a long, TSE-free life, and mostly did not accumulate PrP^{Sc} in their brains (B. Chesebro, Rocky Mountain Labs, Hamilton, Montana). The injection of PrP^{Sc}-negative brain extracts from these mice into further mice again resulted in no clinical disease, over a study period of more than 650 days. But when brain extracts from the latter mice were injected into hamsters, the animals died rapidly. So the infectious agent had silently replicated for several years in mice but maintained full virulence towards hamsters. Given that mouse prions are generally harmless to hamsters, it is hard to understand why — in this instance — they retained their infectivity. It would seem that prion strain characteristics dominate over the amino-acid sequence of the prions from the infected host. It is challenging, but maybe not impossible, to reconcile these data with the protein-only prion hypothesis: maybe the strain-specific properties are really encoded in the tertiary or quaternary structure of PrP^{Sc} rather than in its amino-acid (primary) sequence.

At the genetic level, variations in the human prion gene that protect against the

development of CJD have disseminated much more efficiently than non-protective variations throughout human populations worldwide (J. Collinge, Inst. Neurology, London)⁴. This provides a compelling case that these protective changes were 'selected for' during the course of evolution — but why would they have been necessary? Collinge suggested that they protected against cannibalism-transmitted prion diseases. He derived the disturbing conclusion that cannibalism was once commonplace among our ancestors, and that prion diseases such as kuru — once a prime cause of death in New Guinea tribes that practised cannibalism — ravaged human populations in the distant past.

On a different note, it was suggested that the concept of prions as proteins that exist in two or more conformations, and replicate by causing other such proteins to change shape, applies not just to those proteins implicated in TSEs (R. Wickner, NIH, Bethesda, Maryland)⁵. Thus, in yeast, heritable traits can be attributable to prions under the following conditions: if they can disappear and later reappear (in the same or a later generation); if their propagation depends on the presence of the gene encoding the protein in question; and if their spontaneous frequency increases upon overexpression of that protein. If

one accepts this definition, self-activating enzyme precursors (zymogens) might fall into this category. For instance, yeast protease B will self-activate by limited proteolysis (T. Roberts, NIH, Bethesda, Maryland, and R. Wickner). As the activated state is stable and transmissible by transferring the cytosol from one cell to another, it could be regarded as a prion.

Moving into the clinic, we find improvements in the diagnosis of prion diseases. The conformation-dependent immunoassay allows sensitive diagnosis by exploiting the fact that parts of the prion protein become inaccessible to antibodies during conversion to PrP^{Sc} (S. Prusiner, Univ. California, San Francisco). This led to the discovery that, in certain circumstances, the disease-associated prion protein is conformationally changed but is not protease-resistant (J. Safar, Univ. California, San Francisco)⁶. So, conformational assays might be inherently more sensitive than the venerable protease-based tests used currently.

But the most startling diagnostic development is the advent of a quantitative, sensitive assay that measures the ability of samples to infect prion-susceptible cells (C. Weissmann, Imperial College, London). The test is inexpensive and can be performed in days, as opposed to conventional infectivity assays in

Animal behaviour

Homing is a breeze for sea turtles

Charles Darwin was one of the first biologists to wonder how migrating green turtles find Ascension Island, a 10-kilometre-long speck of land in the vast mid-Atlantic Ocean (*Nature* **7**, 360; 1873). These turtles arrive at Ascension Island between December and March, having travelled a daunting 2,200 kilometres or so eastwards from their feeding grounds off Brazil. Graeme C. Hays and colleagues now suggest a role for wind-borne information in this remarkable island-finding ability (*Proc. R. Soc. Lond. B* doi:10.1098/rspb.2002.2308; 2003).

The navigational feats of marine turtles make them good subjects for researchers interested in migratory behaviour. Early theories proposed that green turtles (*Chelonia mydas*; pictured) pinpoint Ascension Island by using water-borne odours, carried towards Brazil by the South Atlantic Equatorial Current. More recent laboratory studies have shown that hatchling loggerhead turtles can detect the inclination and intensity of the Earth's magnetic field, perhaps

using these features as 'magnetic markers'. However, these promising ideas have yet to be tested on sea turtles in their natural environment.

Hays *et al.* have now performed just such a test of another theory: that green turtles use wind-borne cues to find their way. The authors found six female turtles that had just nested on Ascension Island, and then 'displaced' them. During the nesting period, persistent trade-winds blow from the southeast. So three of the turtles were transported 50 kilometres upwind of the island, and the other three 50 kilometres downwind. The animals' movements were then tracked by satellite. Female *C. mydas* come ashore to lay eggs several times in a season, so they had a strong urge to return to lay their remaining eggs.

Incredibly, Hays *et al.* found that the three downwind-displaced turtles returned to the island within 1, 2 and 4 days. Two of those moved upwind eventually returned after 10 and 27 days — but only after passing downwind of the island. A third



upwind-displaced turtle was unable to locate Ascension Island after 59 days, and was heading back to Brazil when satellite transmissions ceased.

It seems, then, that wind-borne cues emanating from the island — perhaps odours or sounds — might direct turtles. The authors concede that such cues will not work accurately over very long distances, so might not on their own account for island finding. But perhaps a combination of mechanisms is at work, with a geomagnetic sense aiding navigation over large scales (hundreds to thousands of

kilometres), and smell or hearing operating on scales of tens of kilometres. Other long-distance movers might also use such a scale-dependent battery of sensory 'channels' for migration. Identifying the role of particular channels, and when they kick in, will be a challenge, but one that will surely benefit from comparisons of turtles with other vertebrates. **David W. Sims** David W. Sims is at the Marine Biological Association, The Laboratory, Citadel Hill, Plymouth PL1 2PB, UK. e-mail: dws@mba.ac.uk

which scores of animals must be observed for months or even years. For now, however, only mouse prionologists will profit, as the test has not been adapted to other species.

Little progress was reported on the therapeutic side. The antimalarial drug quinacrine can cure prion-infected cells⁷, but hard evidence for any anti-prion effects *in vivo* is still lacking⁸. Instead, quinacrine-treated CJD patients suffer severe liver damage (N. Streichenberger, Univ. Hospital, Lyons)⁹. Stimulation of a specific type of signalling pathway in the innate immune system delays prion diseases, perhaps by eliciting the production of anti-PrP antibodies (H. Kretzschmar, Univ. Munich)¹⁰, or perhaps because long-lasting stimulation might lead to immune disruption (M. Heikenwalder), which has been documented to slow the progression of prion diseases. Finally, the anti-prion properties of soluble PrP derivatives might merit further study (A. Aguzzi)¹¹.

It is clear that exciting times lie ahead in this field. Prion diseases are far from understood, and there are many bones of contention. We could not escape the exhilarating feeling that, more than in other areas of biology, fundamental discoveries are yet to be made. ■

Adriano Aguzzi and Mathias Heikenwalder are at the Institute of Neuropathology, University Hospital of Zurich, CH-8091 Zurich, Switzerland.
e-mail: adriano.aguzzi@usz.ch

1. Ma, J. & Lindquist, S. *Science* **298**, 1785–1788 (2002).
2. Ma, J., Wollmann, R. & Lindquist, S. *Science* **298**, 1781–1785 (2002).
3. Drisaldi, B. *et al.* *J. Biol. Chem.* doi:10.1074/jbc.M213247200 (26 March 2003).
4. Mead, S. *et al.* *Science* **300**, 640–643 (2003).
5. Wickner, R. B. *Science* **264**, 566–569 (1994).
6. Safar, J. *et al.* *Nature Med.* **4**, 1157–1165 (1998).
7. Korth, C., May, B. C., Cohen, F. E. & Prusiner, S. B. *Proc. Natl Acad. Sci. USA* **98**, 9836–9841 (2001).
8. Collins, S. J. *et al.* *Ann. Neurol.* **52**, 503–506 (2002).
9. Scoazec, J. Y. *et al.* *Ann. Neurol.* **53**, 546–547 (2003).
10. Sethi, S., Lipford, G., Wagner, H. & Kretzschmar, H. *Lancet* **360**, 229–230 (2002).
11. Meier, P. *et al.* *Cell* **113**, 49–60 (2003).

Cell biology

Patches for wounded muscle

Juliet A. Ellis

The stresses and strains imposed on certain cells mean that their membranes require constant repair. Study of the damage that affects muscle membranes reveals a new component of the repair process.

Cell membranes in tissues such as skin, gut and muscle are routinely exposed to mechanical damage, which can produce holes in them. When that damage is not repaired, the consequences can be severe — often resulting in cell death — and may contribute to the development of the muscle degenerative diseases termed muscular dystrophies. From a combination of observations on human muscular dystrophy patients and experiments with mice, Bansal *et al.*¹ (page 168 of this issue) now report that a protein called dysferlin is a component of the mechanism for resealing the holes, and thus healing the muscle membrane.

Membrane resealing is generally carried out by a mechanism that resembles the calcium-regulated release of vesicles from a cell (exocytosis). The repair pathway is initiated by an influx of calcium through a wound, resulting in an increase in calcium levels at the site of injury. This, in turn, triggers the accumulation of vesicles, which fuse with one another and then with the plasma membrane, within the injury. A 'patch' is thereby added across the wounded area, resealing the plasma membrane². The entire process — which remains largely mysterious — takes between ten and thirty seconds.

Specific participants include members of

the SNARE and SNAP family of proteins, which are associated with vesicle fusion in nerve transmission. Among them is a protein called synaptotagmin, which is thought to act as a calcium sensor through its possession of two C2 domains. This feature means that it can bind phospholipids — which are the main components of membranes — in a calcium-dependent manner. The protein investigated by Bansal *et al.*, dysferlin, is found in the muscle plasma membrane (sarcolemma) and in cytoplasmic vesicles, and its participation in membrane repair is all the more thought-provoking given its association with muscle degeneration.

Mutations in dysferlin cause two conditions — limb-girdle muscular dystrophy type 2B (LGMD2B) and Miyoshi myopathy (MM). Both are muscular dystrophies, which are all characterized by skeletal-muscle wasting and weakness (see Box 1). Many of the genes implicated in this group of diseases encode proteins that make up the 'dystrophin-glycoprotein complex', which straddles the sarcolemma and crosslinks structural components inside the cell with the extracellular environment³. The complex stabilizes muscle fibres by acting as a shock absorber for the forces of contraction and relaxation to which they are continually subjected. Mutations in any component of the dystrophin-glycoprotein complex lead to a secondary loss of other components, and to the eventual breakdown of the sarcolemma.

Muscle fibres isolated from LGMD2B/MM patients exhibit disruptions to the sarcolemma that are characterized by the appearance of clusters of vesicles beneath the damage⁴. Bansal and colleagues generated mice that don't produce dysferlin (dysferlin-null mice), and these animals developed symptoms of progressive muscular dystrophy similar to those seen in human patients. But neither the dysferlin-null mice nor the LGMD2B patients showed any evidence of malfunction of the dystrophin-glycoprotein complex, or any sign of sarcolemma instability. So, given that dysferlin and several other proteins implicated in muscular dystrophy

Box 1 Types of muscular dystrophy

The muscular dystrophies include about 50 diseases that affect the muscles, particularly those of the arms and legs. In these diseases, the muscles weaken with age: affected people find it increasingly difficult to walk, many becoming wheelchair-bound in their mid-teens. Some dystrophies are also associated with heart problems, which can be life-threatening. The defects are all due to a breakdown in the integrity of the cells that make up muscle fibres,

which gradually wither away. The most common and severe muscular dystrophies are Duchenne, which mainly affects young boys, and a form of myotonic dystrophy that develops in adults. The Muscular Dystrophy Campaign, UK, estimates that, taken as a group, worldwide these conditions may affect as many as 1 in 2,000 people born (most sufferers die young, and so the incidence in the adult population is less).

The problem in Duchenne

muscular dystrophy is due to a fault in a protein, dystrophin, that lies just under the cell surface. Its prime function is to strengthen the muscle cell by acting like a kind of scaffolding which, in a complex with various glycoproteins, connects the structural components inside the cell with the extracellular environment. Other dystrophies are associated with proteins, such as laminin- α 2 and the sarcoglycans, that link to dystrophin, assisting it in its scaffolding role.

Another form of the disease arises from defects in proteins in other parts of the cell that also contribute to the cell's integrity. Examples are the nuclear proteins emerin and lamin A/C, and the cell-surface proteins caveolin-3 and dysferlin. At present, there is no cure for the muscular dystrophies. Most research is centred on developing animal and cell models that mimic the dystrophy symptoms observed in patients, as well as on the design of gene-therapy trials. **J.A.E.**

are not components of the dystrophin–glycoprotein complex, it was reasonable to conclude that other mechanisms could generate some of the features of muscular dystrophies.

The amino-acid sequence of dysferlin is 28% identical to that of Fer-1 protein found in the nematode worm *Caenorhabditis elegans*. In *C. elegans*, Fer-1 is proposed to mediate vesicle fusion to the plasma membrane of one cell type⁵. It is part of a growing family of 'ferlins', which include myoferlin and otoferlin, and is characterized by having six C2 domains (although only the first of them confers a calcium-dependent, phospholipid-binding ability). One LGMD2B patient has been found to have a point mutation in this C2 domain, reducing the domain's phospholipid-binding ability⁶. Myoferlin⁷ is also a sarcolemma protein, and its activity is increased in mice — *mdx* mice — that lack dystrophin, while otoferlin⁸ is associated with vesicle fusion in the inner ear.

Given this background, Bansal *et al.* reasoned that dysferlin could well be involved in calcium-dependent membrane repair, and they set out to test this hypothesis. To do this, they exposed isolated muscle fibres from normal — wild-type — mice to laser-induced membrane damage. Vesicles that stained positive for dysferlin accumulated beneath the sites of injury. To see if these vesicles could participate in membrane repair, Bansal *et al.* isolated muscle fibres from wild-type, *mdx* and dysferlin-null mice, exposed them to laser-induced membrane damage, and then timed how quickly their sarcolemmas resealed themselves in the presence and absence of external calcium. In both wild-type and *mdx* mice, sarcolemmas were resealed within a minute of being damaged, but only when calcium was present. The membranes of muscle fibres in dysferlin-null mice failed to resealed in any circumstances, confirming the role of dysferlin in membrane repair. So the damage seen in muscular dystrophies can indeed arise from a mechanism that is not associated with the dystrophin–glycoprotein complex.

Another group has reported⁹ that vesicles containing dysferlin undergo exocytosis in 57% of muscle fibres examined from patients with defective dysferlin, suggesting that dysferlin-containing vesicles may be part of an intracellular transport pathway. If myoferlin or otoferlin (or both) are also present in these vesicles, it seems likely that other ferlin family members function in membrane repair. If so, defects in proteins other than dysferlin would be implicated in producing certain hallmarks of muscular dystrophy by preventing efficient wound-healing.

Interestingly, dysferlin has also been reported^{10,11} to associate with the calcium-dependent protease calpain 3 and the sarcolemma protein caveolin-3. This is significant because defects in both of these proteins

cause a type of muscular dystrophy, although neither protein belongs to the signalling routes implicated in the membrane repair process discussed here. Instead, modulation of calpain activity contributes to muscular dystrophies by disrupting cell-regulatory mechanisms, whereas caveolin-3 organizes lipid and protein constituents in the plasma membrane as part of a vesicular transport mechanism. Evidently, there may be yet more pathways to be discovered that, when disrupted in some way, ultimately lead to sarcolemma fragility.

Juliet A. Ellis is in the Randall Centre for Molecular Mechanisms of Cell Function, King's College, New Hunt's House, Guy's Campus, London SE1 1UL, UK.

e-mail: juliet.ellis@kcl.ac.uk

1. Bansal, D. *et al.* *Nature* **423**, 168–172 (2003).
2. McNeil, P. L. & Terasaki, M. *Nature Cell Biol.* **3**, E124–E129 (2001).
3. Durbeej, M. & Campbell, K. P. *Curr. Opin. Genet. Dev.* **12**, 349–361 (2002).
4. Piccolo, F., Moore, S. A., Ford, G. C. & Campbell, K. P. *Ann. Neurol.* **48**, 902–912 (2000).
5. Bashir, R. *et al.* *Nature Genet.* **20**, 37–42 (1998).
6. Davis, D. B., Doherty, K. R., Delmonte, A. J. & MacNally, E. M. *J. Biol. Chem.* **277**, 22883–22888 (2002).
7. Davis, D. B., Delmonte, A. J., Ly, C. T. & MacNally, E. M. *Hum. Mol. Genet.* **9**, 217–226 (2000).
8. Yasunaga, S. *et al.* *Nature Genet.* **4**, 363–369 (1999).
9. Selcen, D., Stilling, G. & Engel, A. G. *Neurology* **56**, 1472–1481 (2001).
10. Anderson, L. V. B. *et al.* *Neuromusc. Disord.* **10**, 553–559 (2000).
11. Matsuda, C. *et al.* *Hum. Mol. Genet.* **10**, 1761–1766 (2001).

Complex fluids

Spread the word about nanofluids

Manoj K. Chaudhury

Liquid spreads and wets a surface, but wetting behaviour changes if the liquid contains nanoparticles. New experiments point to an explanation and a way to create effective detergents for cleaning oil from a surface.

The wetting of solid surfaces by liquids is usually described in terms of the surface and interfacial tensions at the 'three-phase contact line' — where solid surface, liquid and the air above meet. But this simple situation changes drastically when the liquid is a dispersion of nanoparticles. On page 156 of this issue, Darsh Wasan and Alex Nikolov¹ relate experiments and theoretical modelling that explain why this is. Ordering of

the nanoparticles, they say, creates extra, osmotic pressure — an effect that could be exploited in the detergents used to clean up after oil spills.

The surface of a liquid is unique in the sense that it can bend inwards or outwards, maintaining a capillary pressure within a liquid drop that is lower or higher than the surrounding pressure. Excess pressure (known as the disjoining pressure) can also

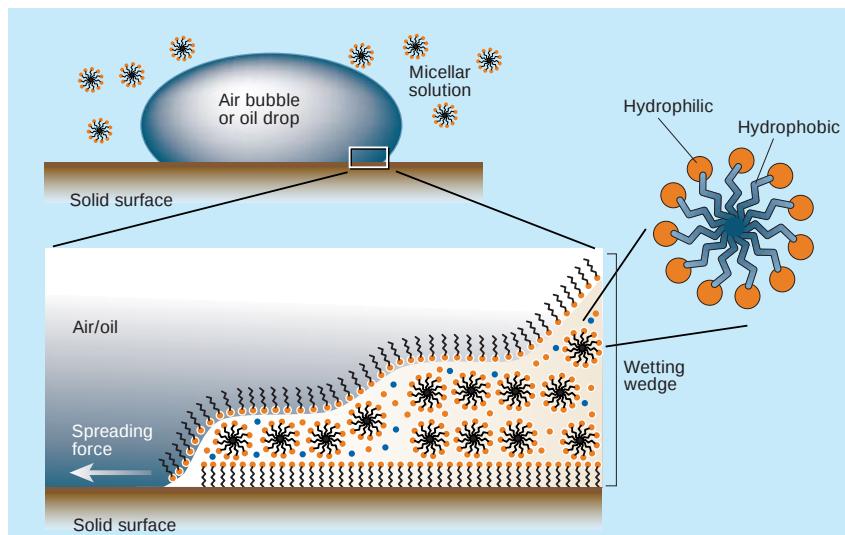


Figure 1 Wet cleaning. Wasan and Nikolov¹ have investigated the wetting behaviour of a solution of nanoparticles — 5-nm-diameter micelles with a hydrophilic surface and hydrophobic core. By introducing air bubbles or oil drops on a glass surface in the micellar solution, they observed the 'wetting wedge' of solution in the confined space between the drop or bubble and the solid surface. It seems that the micelles organize in the wetting wedge wherever the wedge thickness matches an integral multiple of micelle diameters. This induces an osmotic component in the excess pressure directed towards the wedge from the bulk solution. The excess pressure drives further spreading of the solution and can be large enough to separate an oil drop from the solid surface.

develop in a thin liquid film (less than 10 nm thick) when its two surfaces attract or repel each other through van der Waals and electrostatic forces². The interplay of the capillary and disjoining pressures is critical to the stability of thin curved films, such as the wedge-like region at the edge of a wetting solution, trapped between a solid surface and the air above. If one pressure is higher than the other, liquid flows in or out of the wedge, causing the solution to wet or dewet the surface. Integrating the capillary and disjoining pressures over the liquid volume produces Young's equation for simple liquids, which shows how to relate the angle of the wetting wedge to the surface and interfacial tensions of the three interfaces³.

What happens when the wetting solution is not a simple liquid but is instead a concentrated dispersion of nanoparticles? Wasan and Nikolov¹ looked at the effect of using a dispersion of surfactant micelles — self-assembled particles, about 5 nm in diameter, formed of molecules with a hydrophilic and a hydrophobic end (Fig. 1). Micellar spreading is different from classical flow, as here wetting is driven by the gradient of film tension — that is, the tension acting over the depth of the wetting wedge. This phenomenon is also different from the surfactant-driven spreading of a liquid drop caused by the gradient of the surface tension on an expanding interface.

Uniform-sized micelles are known to form ordered structures^{4–6} in confined spaces, such as in thin soap films, imparting stability to these structures. Wasan and Nikolov¹ have discovered that the micelles also form ordered structures near the three-phase contact line of a drop on a solid surface, which promotes wetting. The authors suggest that the micelles form ordered domains in the confined space of the wetting wedge wherever its thickness is equal to an integral multiple of the diameter of a single micelle (Fig. 1). There are two pronounced effects that result from this ordering. First, because the concentration of micelles is larger in the wedge than in the bulk suspension of the surrounding liquid, an osmotic pressure develops that attempts to separate the two interfaces, increasing the depth of the wedge between them. Second, because the film tension (the integrated value of the disjoining pressure over the width of the wedge) increases towards the vertex of the wedge, this creates an extra driving force for spreading the liquid at the wedge tip.

Wasan and Nikolov have explored these effects experimentally, observing the wetting wedge formed by an air bubble on a glass plate through an aqueous dispersion of 1- μ m-diameter latex spheres (which is similar to a micellar suspension but allows visual inspection). They saw evidence that the latex particles assemble in the wedge in ordered structures: the average distances between

particles followed a damped oscillatory pattern consistent with the prediction of statistical mechanics (Fig. 1b on page 156).

In a second experiment, a concentrated dispersion of 5-nm-diameter micelles was injected over an oil drop resting on a glass plate. Video microscopy showed a rapid invasion of the oil–water interface by the aqueous solution, and, as expected because of the increased disjoining pressure, efficient removal of the oil drop. But adding an electrolyte (in this case, sodium chloride) to the nanoparticle solution unexpectedly caused the oil-removal effect to disappear. Wasan and Nikolov suggest that the presence of the electrolyte caused the micellar diameter to shrink, leading to a decrease in the volume fraction of micelles in the wedge, and thereby decreasing the film tension.

The findings of Wasan and Nikolov¹ are of tremendous significance for many areas of research — from the interactions of liquid drops in complex fluids to the 'superwetting' of concentrated surfactant solutions on hydrophobic surfaces — as well as for applying

cations such as the recovery of spilled oil. But to understand the phenomenon of wetting more fully, future studies will need to explain the roles of solvent structuring and the capillary pressure resulting from the curvature of the liquid meniscus in the stratified region of the wedge. These are important considerations, as the overall film pressure should be monotonic from the bulk solution all the way to the vertex of the wedge, even if the film pressure due to the presence of micelles oscillates. ■

Manoj K. Chaudhury is in the Department of Chemical Engineering, Lehigh University, 111 Research Drive, Iacocca Hall, Bethlehem, Pennsylvania 18015, USA.
e-mail: mkc4@lehigh.edu

1. Wasan, D. T. & Nikolov, A. D. *Nature* **423**, 156–159 (2003).
2. de Gennes, P. G. *Rev. Mod. Phys.* **57**, 827–863 (1985).
3. Hirasaki, G. J. *SPE Formation Evaluation* 217–226 (June 1991).
4. Pollard, M. L. & Radke, C. J. *J. Chem. Phys.* **101**, 6979–6991 (1994).
5. Marinova, K. G., Gurkov, T. D., Dimitrova, T. D., Alargova, R. G. & Smith, D. *Langmuir* **14**, 2011–2019 (1998).
6. Wasan, D. T. & Nikolov, A. D. in *Supramolecular Structure in Confined Geometries* (eds Manne, S. & Warr, G.) ACS Symp. Ser. **736**, 40–53 (Oxford Univ. Press, New York, 1999).

Computational biology

Biosensor design

William F. DeGrado

A series of bacterial receptor proteins have been 'redesigned' by computer so that they bind molecules that are quite different from their natural ligands. The approach might be useful for designing catalytic proteins.

On page 185 of this issue, Looger and colleagues¹ describe a powerful computational method for designing proteins that can detect small molecules. The authors have tested their approach on compounds such as trinitrotoluene (TNT) and the neurotransmitter serotonin, and their findings might see applications in, for instance, medicine and biotechnology.

Organisms use a broad repertoire of small-molecule-binding proteins — such as receptors and antibodies — to mediate cell-to-cell communication, signalling, and protection against pathogens. Binding proteins such as antibodies are also used widely to diagnose and treat diseases, and as molecular sensors. Until now, such proteins have been 'developed' *in vivo*, or by *in vitro* methods^{2,3}. The *in vitro* methods require the construction of large 'libraries' of proteins containing diverse amino-acid sequences, coupled with an efficient strategy for selecting and continuously evolving those proteins that can bind the target molecule (the ligand) tightly. This approach is inevitably highly time-consuming. Looger *et al.*¹ have now accomplished the tasks of library construction and screening much more rapidly — in a computer.

The authors started with a series of

periplasmic binding proteins⁴ (PBPs) from the bacterium *Escherichia coli*. These Venus-flytrap-like receptor proteins have two structural domains that clamp together when they engage specific nutrient molecules. On binding their ligand they send signals into the cell, ultimately leading to the activation of appropriate genes. Looger *et al.* used a computational method to change the specificities of the PBPs entirely, so that they bind TNT, serotonin or the sugar L-lactate instead of their usual nutrient (Fig. 1).

The computational design process was initiated by placing a 'virtual' molecule of interest in the binding site of a virtual PBP receptor. The program then sequentially mutated receptor amino acids involved in binding the ligand, searching for sequences that form a surface complementary to the ligand. Typically, 12 to 18 amino acids were mutated, so this step alone created up to 10²³ possible sequences — significantly more than can be screened *in vitro*. However, when approaching this problem computationally, not only must the amino-acid sequence of the receptor be specified, but also the orientation of the ligand, as well as the various conformations that might be adopted by the side chains of the mutated amino acids. These requirements greatly expanded the

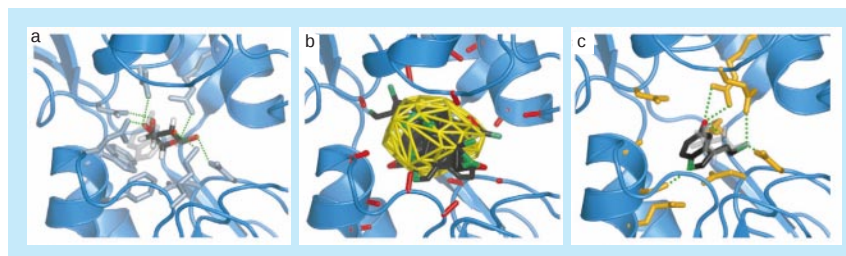


Figure 1 Computational design of binding proteins. Looger *et al.*¹ have converted bacterial periplasmic binding proteins (PBPs) into receptors that bind serotonin, L-lactate or trinitrotoluene (TNT). Steps in the design of a serotonin receptor are shown. **a**, Close-up of a wild-type PBP that binds the sugar arabinose. The sugar is black; the complementary surface of the receptor is pale blue. A ghostly image of serotonin indicates its position in the redesigned binding pocket. **b**, An intermediate step. Amino acids in the complementary surface have been computationally truncated to alanines (red), leaving a large cavity into which sterically allowed serotonin molecules are docked. These molecules are confined to a convex polygon (yellow), restricting them to roughly the space occupied by arabinose. **c**, The serotonin receptor. The automated design algorithm has identified an orientation of serotonin that has a complementary surface (gold). This was predicted (and shown experimentally) to convert the wild-type protein into a serotonin receptor. The original arabinose is shown as a ghost image. Ligand atoms: carbons, black; nitrogens, green; oxygens, red. Dashed green lines represent hydrogen bonds. (Figure supplied by H. W. Hellinga.)

combinatorial problem to between 10^{53} and 10^{76} permutations.

Finding productive combinations in such an astronomical sea of possibilities requires powerful computational algorithms⁵. Looger *et al.* used an enhanced version of 'dead-end elimination'⁶, a tree-pruning method that considers two-body interactions to eliminate features that cannot belong to the global energy minimum. This strategy has previously been used to great advantage in the automated design of proteins, to determine amino-acid sequences that fold into well-defined three-dimensional structures⁷. Of course, the success of the computed sequence depends on how the energy (or 'fitness') of the interacting groups is calculated. Looger and colleagues' calculations were relatively simple, including terms representing hydrogen bonds, van der Waals interactions, electrostatic interactions and atomic solvation.

Seventeen of the virtually designed receptors were thereby selected for experimental testing. To carry out these experiments, the authors first mutated the real-life PBPs as specified by the computational algorithm. Then, in one set of studies, they modified the mutant PBPs with a dye whose fluorescence-emission intensity changed when the proteins bound the correct ligand. Fluorescence measurements showed that the redesigned receptors bound their new ligands specifically and with affinities ranging from relatively low values to values as high as those typically observed in antibody–ligand complexes. In further experiments, the authors inserted the mutant PBPs into *E. coli*, linking the receptors through a synthetic signalling pathway to the production of a 'reporter' enzyme. These techniques provide the possibility of experimentally enhancing the binding still further, by selection or screening.

Finally, the authors returned to the computer and calibrated the relative contribution (weight) of each term in the energy calculation against the experimental data. To do this, they used a quantitative structure–activity relationship (QSAR) that constructs a linear relationship between the observed affinity and the calculated properties of the designs. In the future, this approach should allow more accurate design of high-affinity binders.

There have been previous successful attempts to design proteins that bind specific peptide ligands. For instance, we computationally designed a protein that bound a specific region (peptide) of the calcium-activated enzyme calcineurin⁸. We did so by introducing the sequence of the peptide onto the crystal structure of a newly designed bundle of three α -helices (a common structural unit of proteins), changing the sequence of one helix, and then moulding the sequence of the remaining two-helix bundle to the shape and electrostatic profile of the peptide. Similarly, computer-aided design was used to change the peptide-binding specificity of a PDZ domain — a protein–protein-interaction motif — by first changing the sequence of the bound peptide and then selecting for mutations in the PDZ domain that would accommodate these changes⁹.

So what makes Looger and colleagues' work different? First, in our earlier study⁸ the backbone of the ligand was held fixed relative to that of the receptor protein. This fixed geometry led to underprediction of the number of possible binding sequences, and severely limited the method to the design of proteins that bind close analogues of the starting ligand. Second, whereas previous studies considered side chains only in their canonical ideal conformations, or in a few variations on these arrangements, Looger

et al. used a more fine-grained sampling of dihedral angles. This approach allowed optimization of hydrogen bonds to the polar groups in the target ligand. Indeed, the highest-affinity ligands came out of the most fine-grained sampling. Finally, Looger *et al.*'s calculations required that all possible hydrogen bonds be satisfied, ensuring that the protein bound specifically to the target.

The authors suggest a variety of possible applications for their methods — for instance, following further development, the TNT receptor might be useful in detecting landmines, and the serotonin receptor in clinical diagnostics.

The approach might also have potential for the design of catalytic proteins, for applications in, for instance, the chemical industry or diagnostics. When an enzyme catalyses a particular reaction, its substrate normally passes through a transition state. Antibodies that bind analogues of the transition state can speed up the reaction, with good selectivity¹⁰ — but the rate enhancements are generally much lower than those seen with the enzyme. A possible explanation for this is that the transition-state analogues are less than perfect mimics of the true transition state. Greater rate accelerations might be achieved by designing proteins that bind a virtual transition state that is an accurate mimic of the true transition state, and whose geometry and charge distribution can be estimated using quantum-mechanical calculations. A related approach has been used to computationally design an enzyme from an inert protein. By targeting a high-energy covalent intermediate, Bolon and Mayo¹¹ designed a protein that catalysed the hydrolysis of 4-nitrophenyl esters (reviewed in ref. 12). Although the rate enhancements were modest, the protein showed enzyme-like saturation kinetics, illustrating the promise of the method. The approach of Looger *et al.*, which does not require the formation of a covalent intermediate, should provide a complementary means of designing catalysts. ■

William F. DeGrado is in the Department of Biochemistry and Biophysics, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6059, USA.

e-mail: wdegrado@mail.med.upenn.edu

- Looger, L. L., Dwyer, M. A., Smith, J. J. & Hellinga, H. W. *Nature* **423**, 185–190 (2003).
- Hoess, R. *Chem. Rev.* **101**, 3205–3218 (2001).
- Dower, W. J. & Mattheakis, L. C. *Curr. Opin. Biotechnol.* **6**, 390–398 (2002).
- Tam, R. & Saier, M. H. Jr *Microbiol. Rev.* **57**, 320–346 (1993).
- Desjarlais, J. R. & Clarke, N. D. *Curr. Opin. Struct. Biol.* **8**, 471–475 (1998).
- Lasters, I., De Maeyer, M. & Desmet, J. *Protein Eng.* **8**, 815–822 (1995).
- Dahiyat, B. I. & Mayo, S. L. *Science* **278**, 82–87 (1997).
- Ghirlanda, G., Lear, J. D., Lombardi, A. & DeGrado, W. F. *J. Mol. Biol.* **281**, 379–391 (1998).
- Reina, J. *et al.* *Nature Struct. Biol.* **9**, 621–627 (2002).
- Wentworth, P. Jr & Janda, K. D. *Cell Biochem. Biophys.* **35**, 63–87 (2001).
- Bolon, D. N. & Mayo, S. *Proc. Natl Acad. Sci. USA* **98**, 14272–14279 (2001).
- Baltzer, L. & Nilsson, J. *Curr. Opin. Biotechnol.* **12**, 355–360 (2001).

Cell biology

Drug resistance lies within

Mol. Biol. Cell doi:10.1091/mbc.E02-11-0704 (2003)

When tumours become drug-resistant they often ramp up the production of so-called multidrug-resistance transporters — thought to thrust the toxin out of the tumour's cells. In fact, these pumps might be working inside the cells instead.

Asha Rajagopal and Sanford M. Simon genetically engineered cells in culture so that they made a fluorescent version of one of a family of pumps, such as multidrug-resistance protein 1 and breast-cancer-resistance protein. They then tracked the fate of chemotherapy drug doxorubicin on the basis of its natural fluorescence. They found that the pumps prevent doxorubicin from reaching its target, the nucleus, and killing cells. Instead, the drug is diverted into small compartments called lysosomes, which break down cellular trash.

The transporters sit in the membrane of lysosomes and probably pump in the drugs, rendering them useless against the cell. The authors say that this could affect the design of treatments aimed at quashing drug-resistance proteins.

Helen Pearson

Environmental science

Shades of China

Glob. Biogeochem. Cycles **17**, 1034–1052 (2003)

An analysis of historical tree coverage in China suggests that the country was a net source of atmospheric carbon well before its industrialization, before becoming a carbon sink in the past decade or so. Attempts at such analyses have previously been stymied by a lack of historical information: the first national inventory of forests was not published until 1983, for example. But by examining these data, and augmenting them with agricultural archives and inferring forest cover from historical accounts and pollen records, R. A. Houghton and J. L. Hackler have been able to produce estimates of land-use change in China between 1700 and 2000.

Pollen records indicate that by 1700 China had already lost between 19% and 48% of its estimated prehistoric forest cover. Continued deforestation led to the loss of a further 180 million hectares between 1700 and 2000, with the concomitant release of some 17–33 billion tonnes of carbon into the atmosphere.

Between the 1950s and 1970s, emissions of carbon due to land-use change peaked at 0.2 to 0.5 billion tonnes annually. But more recently the picture has changed: because of the implementation of nationwide re-forestation programmes,



and a 1998 moratorium on logging, during the 1990s China has on average become a sink for atmospheric carbon.

Tom Clarke

Ecology

News of chews

Funct. Ecol. **17**, 201–212 (2003)

Differences in jaw power might help different bat species coexist, say L. F. Aguirre and colleagues. They looked at the diets and jaw strengths of 23 bat species living on the Bolivian savanna, and found that the diet of each species seems to be limited by the strength of their bites.

Analyses of stomach contents revealed that each species specialized in either fruit or insects. For both fruit and insects, bigger food items tended to be harder. And the maximum size of food item in each species' diet tended to match the strength of its jaws, as estimated by force measurements of its bite. Food hardness probably limits the diet of small bat species; only a few large species are able to crunch their way through the bigger beetles or toughest fruits. Bats are known to chew their food well, so the time and energy taken to process tough foodstuffs are probably an important determinant of what they can eat — and so how different species divide up the food resources in their environment.

John Whitfield

Microscopy

Nanoscience in ribbons

Appl. Phys. Lett. **82**, 2886–2888 (2003)

The sensitivity of an atomic force microscope (AFM) depends on the thin ribbon of material that connects the needle-like probe tip to the body of the device. This cantilever should be thin and stiff, and is typically made of a hard, strong substance

such as silicon nitride, giving it a large spring constant and the ability to detect very small forces. William L. Hughes and Zhong L. Wang claim that it should be possible to make vastly improved AFM cantilevers from 'nanobelts' of metal oxides and sulphides.

Nanobelts are tiny ribbons that form spontaneously when the materials are deposited by high-temperature evaporation of a fine powder — an example of the 'bottom-up' approach to nanofabrication, contrasting with the way that cantilevers are traditionally made by carving them from bigger chunks of material. Wang and co-workers first made nanobelts two years ago. The ribbons are typically 20–1,000 nm wide — up to 1,800 times smaller than conventional cantilevers.

Now Hughes and Wang have developed methods for cutting the ribbons into shorter sections: for example, using the electron beam of a transmission electron microscope, or chiselling through a nanobelt using a conventional AFM tip. They then use an AFM to pick up a section of the ribbon and position it on a silicon wafer. If a fine tip could be attached to the end of a ribbon, it would serve as a peerless cantilever: the increase in force sensitivity might realize such alluring goals as single-atom magnetic-resonance spectroscopy.

Philip Ball

Animal behaviour

A dog's dinner

Curr. Biol. **13**, 763–766 (2003)

Dogs readily understand human gestures that direct them to food. Pointing, a nod of the head, or simply gazing in the direction of a hidden treat can get them going. How did dogs obtain this skill? From their comparison of wolves and dogs, Ádám Miklósi *et al.* propose that it has to do with centuries of domestication.

Miklósi *et al.* discovered that teaching wolves to find a piece of meat hidden in a pot wasn't easy, even when the wolves were reared as if they were dogs. Touching the pot containing the meat, or pointing at it from nearby, helped a little, but only after many trials did the wolves respond to pointing from a distance.

In order to spot direction in a distant gesture, an animal must look at the trainer, and the wolves seemed to avoid just that. When confronted with an insoluble task — retrieving meat from a locked pot, for example — the dogs immediately looked back at their owners, and gazed at their face as if to ask for help. But the wolves never looked back.

Domestication may have selected for the tendency of dogs to look at the human face in situations of uncertainty. Miklósi *et al.* suggest that this simple change in behaviour may have opened the door to our long and productive relationship with dogs.

Marie-Thérèse Heemels

A fly in the biogeographic ointment

A tiny fossil provides a clue as to how things were in Antarctica millions of years ago.

We have discovered a fossil of a higher fly (Diptera: Cyclorrhapha) from Antarctica, a finding that goes against the long-held belief that the continent was never inhabited by these insects¹. The fly must either have colonized Antarctica during a warm interval in the Neogene epoch, between 3 million and 17 million years ago, or it was an original member of the Gondwana fauna that survived in Antarctica for tens of millions of years before becoming extinct.

In a biogeographic study of the flies of New Zealand, Australia, southern South America and the sub-Antarctic islands², the great cladist Willy Hennig concluded that although more primitive flies might have inhabited the Mesozoic and Tertiary forests of Antarctica, there was no evidence that the higher flies had ever done so. His analysis suggested that the more derived clades of cyclorrhaphans had evolved on Laurasia after Pangaea had broken up. He postulated that the existing cyclorrhaphans of the southern continents had been derived from different northern stocks. No new evidence³ has been available until now to test his hypothesis.

The fossil has several features that enable us to identify it with confidence as a cyclorrhaphan puparium (Fig. 1). These include a single pair of round spiracles, an integument with circular patterning that reflects the way in which chitin was secreted, spines on the ventral welt (Fig. 1b) and the ecdysial scar from the second-instar stage of development. We estimate that the specimen is from a puparium that was about 5.0–7.5 mm in length. The characteristics of the fossil enable it to be assigned to one of the more highly derived clades (Schizophora) of the Cyclorrhapha, but we are unable to assign it to a family.

The fossil is from a siltstone collected by one of us (A.C.A.) from the Meyer Desert Formation, which outcrops on the margins of the Beardmore Glacier at latitude 85° S, about 500 km from the South Pole. A mid-Pliocene age⁴ has been assigned to the formation, although some consider an older minimum age of mid-Miocene more likely⁵. The siltstone was deposited on the margins of a glacier at the head of a wide fjord. The landscape was sufficiently stable to contain a lake, ice-free for long enough during the summers to support algae, freshwater molluscs⁶ and a species of fish. On the better-drained morainic soils, a shrub and herb tundra with cushion plants, *Ranunculus* (buttercups) and patches of dwarf *Nothofagus* (southern beech)⁷

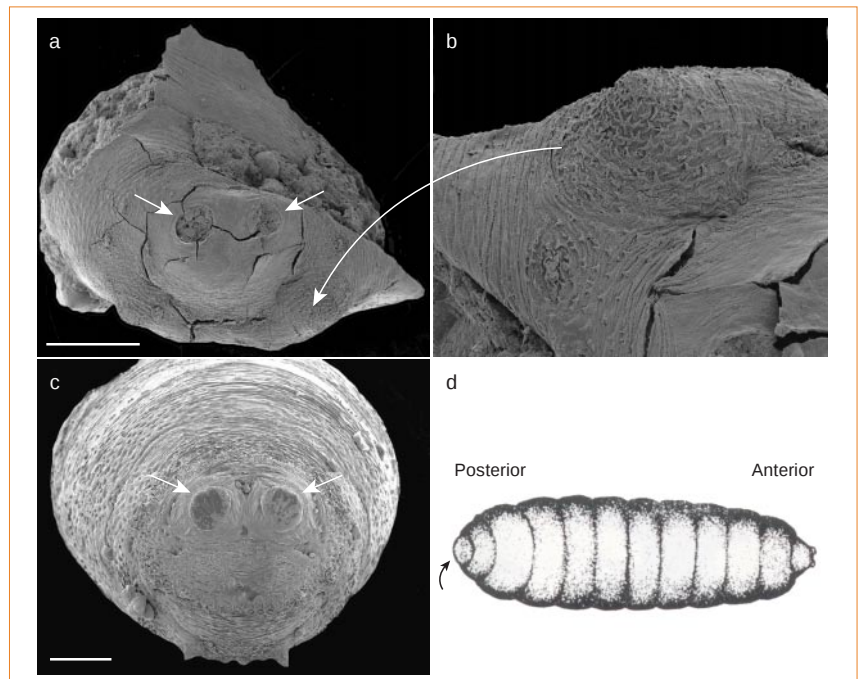


Figure 1 Scanning electron micrographs of the Antarctic fly fossil puparium compared with a modern puparium. The puparium is the hardened integument of the 3rd (last) instar (maggot) from which the adult fly emerges. **a**, Posterior view of the fossil puparium, which is partly crushed and has silt grains embedded in the chitin. The posterior spiracles are indicated by the pair of white arrows. Scale bar, 0.5 mm. **b**, Enlarged view of a ventral cuticular welt on the fossil, showing the presence of spines. **c**, Posterior segment of a modern puparium of a blowfly (*Cochliomyia macellaria*; Fabricius: Calliphoridae). Spiracles are indicated by arrows. Scale bar, 0.5 mm. **d**, Drawing of a puparium to show the location (arrow) of the posterior spiracles in both the fossil (**a**) and a modern specimen (**c**).

was inhabited by listroderine weevils⁸ and other insects.

By the Neogene, Gondwana had fragmented and even the closest land in South America was separated from Antarctica by 1,000 km of ocean. Pollen analysis of cores from the Ross Sea indicates that an impoverished tundra vegetation existed in the Ross Sea area from the Oligocene to the Miocene, and possibly the Pliocene⁹. *Nothofagus* from the Meyer Desert Formation was probably a descendant from the early Cainozoic Gondwana forest, rather than a colonizer during the Neogene⁷. Climatic cooling and the formation of ice sheets, associated with either the opening of the Drake Passage and the formation of the Circumpolar Current between 34 million and 22 million years ago¹⁰ or a decline in atmospheric CO₂ levels¹¹, could ultimately be the cause of the extinction of most Antarctic terrestrial organisms.

In proposing that the Cyclorrhapha had never inhabited Antarctica, Hennig was astute enough to add the caveat “in the absence of a fossil record”. The critical question now is whether the fossil was from a South American population that colonized

Antarctica during an especially warm period during the Neogene or from a relict Gondwana population that had inhabited tundra coastal habitats for millions of years in isolation in Antarctica. If better-preserved fossils are found in Antarctica, a radical revision of the age and place of origin of the Cyclorrhapha might one day be necessary. Until then, dipterists need to re-examine the taxonomy of southern cyclorrhaphan fly taxa to see whether any trans-Antarctic relationships among the more derived clades have been overlooked.

Allan C. Ashworth*, F. Christian Thompson†

*Department of Geosciences, North Dakota State University, Fargo, North Dakota 58105-5517, USA
e-mail: allan.ashworth@ndsu.nodak.edu

†USDA Systematic Entomology Laboratory, Smithsonian Institution, Washington DC 20560, USA

1. Coney, P. & Block, W. *Eur. J. Entomol.* **93**, 1–13 (1996).
2. Wygodzinsky, P. *Pacif. Insects Monogr.* **9**, 1–81 (1966).
3. Amorim, D. de S. & Silva, V. C. *Ann. Soc. Entomol. France* **38**, 177–199 (2002).
4. Webb, P.-N., Harwood, D. M., Mabin, M. G. C. & McKelvey, B. C. *Mar. Micropaleontol.* **27**, 273–297 (1996).
5. Marchant, D. R. *et al. Geol. Soc. Am. Bull.* **114**, 718–730 (2002).
6. Ashworth, A. C. & Preece, R. C. *J. Mollusc. Stud.* **69**, 89–92 (2003).

7. Hill, R. S., Harwood, D. M. & Webb, P.-N. *Rev. Palaeobot. Palynol.* **94**, 11–24 (1996).
8. Ashworth, A. C. & Kuschel, G. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **191**, 191–202 (2003).

Electronic paper

Flexible active-matrix electronic ink display

Ultrathin, flexible electronic displays that look like print on paper are of great interest^{1–4} for application in wearable computer screens, electronic newspapers and smart identity cards. Here we realize the fabrication of such a display on a bendable active-matrix-array sheet. The display is less than 0.3 mm thick, has high pixel density (160 pixels × 240 pixels) and resolution (96 pixels per inch), and can be bent to a radius of curvature of 1.5 cm

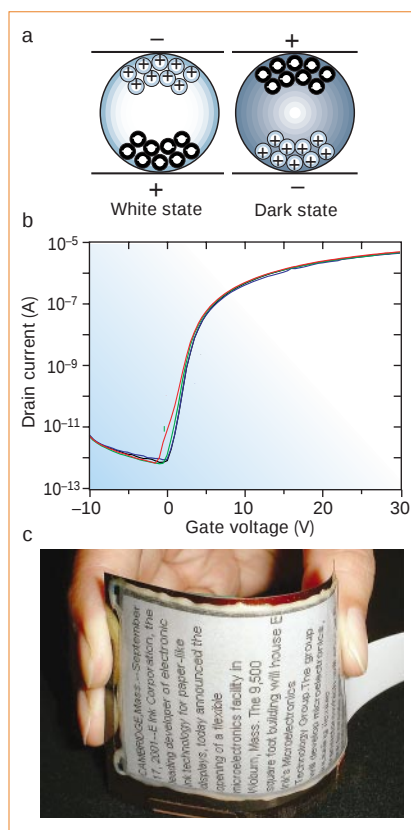


Figure 1 Flexible active-matrix electronic-ink displays. **a**, Operating principle of electronic ink. The relative movement of negatively charged black and positively charged white particles inside their microcapsules is controlled by the direction of the applied voltage. **b**, A backplane thin-film transistor measured *in situ* under compressive stress. The transistor is bent to three different radii of curvature: green, 2.0 cm (0.19% strain); blue, 1.3 cm (0.29% strain); and red, 1.0 cm (0.38% strain). The thin-film transistor has identical characteristics when measured without bending (black curve) and at a radius of curvature of 2.0 cm; degradation is minimal even at 1.0 cm. Results were similar under tensile stress. **c**, Text image shown on a bent display whose resolution is 96 d.p.i. and which has a white-state reflectance of 43% and a contrast ratio of 8.5:1.

9. Askin, R. A. & Raine, J. J. *Terra Antarctica* **7**, 493–501 (2000).
10. Zachos, J. et al. *Science* **292**, 686–693 (2001).
11. DeConto, R. M. & Pollard, D. *Nature* **421**, 245–249 (2003).

Competing financial interests: declared none.

without any degradation in contrast. This use of electronic ink technology on such an ultrathin, flexible substrate should greatly extend the range of display applications.

Thin (0.4-mm) but inflexible liquid-crystal displays have been made on plastic by using a diode-matrix array⁵ and an amorphous-silicon, thin-film transistor (TFT), active-matrix array⁶. To create a flexible display, we used a TFT array (backplane) with microencapsulated electrophoretic material (electronic ink)⁷, which consists of millions of microcapsules containing charged pigment particles in a clear fluid. A negative voltage applied to the top surface causes the positive white particles to move to the top of the capsule and the surface to appear white; reversing the electric field causes the negative black particles to appear at the top surface and create a dark spot (Fig. 1a).

We used a 75-μm-thick steel-foil substrate to build the TFT backplane because steel foil is lightweight, mechanically stable and compatible with existing fabrication processes for active-matrix liquid-crystal displays^{8,9}. Before the array fabrication, an insulating layer was applied onto the foil to render the substrate passive. The amorphous-silicon TFTs were made in the bottom-gate, back-channel etch configuration. The gate and source/drain metal were deposited by sputtering. A ductile composite of aluminium and refractory metal was used for the gate metal to enhance the backplane's flexibility.

Silicon nitride, amorphous silicon and a doped amorphous-silicon layer were deposited as the gate insulator, the channel and the contact layer, respectively, by plasma-enhanced chemical-vapour deposition. The metal, semiconductor and insulator layers were patterned by photolithography. The display was made by laminating a sheet of electronic ink onto the backplane. The electronic ink consists of a layer of electrophoretic microcapsules and a polymer binder, coated onto a polyester/indium–tin oxide (common electrode) sheet. The total display thickness is less than 0.3 mm.

A typical TFT has a threshold voltage of 4.0 volts and a linear mobility of 0.50 cm² V⁻¹ s⁻¹. The drain off current is about 1.0 pA at 10 V drain voltage. The current on/off ratio is 5 × 10⁶, which is sufficient for high-resolution displays. The TFT performance does not degrade after first being bent for 120 seconds around a cylinder that is 2 mm in radius (1.9% strain) and then released.

We also measured TFTs *in situ* under compressive stress at three radii of curva-

ture (Fig. 1b). Because steel has a large Young's modulus, our selection of a thin substrate decreases the distance of the TFT circuit from the display neutral plane¹⁰, reducing the in-plane strain of the circuit. As a result, the display can be repeatedly bent 20 times to a radius of curvature of 1.5 cm without any degradation.

Bias temperature stress on the TFT backplane was performed at a gate voltage of 25 V and up to 80 °C. The results indicate that the flexible backplane has a threshold voltage shift comparable to that of conventional glass TFT backplanes in laptop computers, and possibly a similar reliability and lifetime. The row electrode is driven between 0 and 24 V, and the column electrode is driven between 0 and 20 V.

Figure 1c shows the bent display of a text image of 96 d.p.i. resolution; the display has a viewing angle of almost 180°. The ink-switching speed is 250 ms, which is sufficient for electronic paper. For wearable computers, a reduction to 15 ms would be required for video-rate switching; in addition, the substrate thickness would need to be reduced for foldable displays. We suggest that electronic ink combined with flexible amorphous-silicon active-matrix backplanes will provide a viable pathway to 'e-paper' and wearable computer screens.

Y. Chen, J. Au, P. Kazlas, A. Ritenour, H. Gates, M. McCreary

E Ink Corporation, 733 Concord Avenue, Cambridge, Massachusetts 02138, USA
e-mail: yuc@princeton.edu

1. Huitema, H. E. A. et al. *Nature* **414**, 599 (2001).
2. Rogers, J. A. et al. *Proc. Natl Acad. Sci. USA* **98**, 4835–4840 (2001).
3. Kane, M. G. et al. *IEEE Electron Dev. Lett.* **21**, 534–536 (2000).
4. Sirringhaus, H., Tessler, N. & Friend, R. H. *Science* **280**, 1741–1744 (1998).
5. Baeuerle, R., Baumbach, J., Leuder, E. & Siegfried, J. *SID 99 Tech. Dig.* 14–17 (1999).
6. Polach, S., Randler, M., Bahnmüller, F. & Lueder, E. *IDW 00 Dig.* 203–206 (2000).
7. Comiskey, B., Albert, J. D., Yoshizawa, H. & Jacobson, J. *Nature* **394**, 253–255 (1998).
8. Theiss, S. & Wagner, S. *Mater. Res. Soc. Symp. Proc.* **424**, 65–70 (1997).
9. Chen, Y., Denis, K., Kazlas, P. & Drzaic, P. *SID 01 Tech. Dig.* 157–159 (2001).
10. Suo, Z., Ma, E. Y., Gleskova, H. & Wagner, S. *Appl. Phys. Lett.* **74**, 1177–1179 (1999).

Competing financial interests: declared none.

Comparative genomics

Insecticide resistance in mosquito vectors

Resistance to insecticides among mosquitoes that act as vectors for malaria (*Anopheles gambiae*) and West Nile virus (*Culex pipiens*) emerged more than 25 years ago in Africa, America and Europe; this resistance is frequently due to a loss of sensitivity of the insect's acetylcholinesterase enzyme to organophosphates and

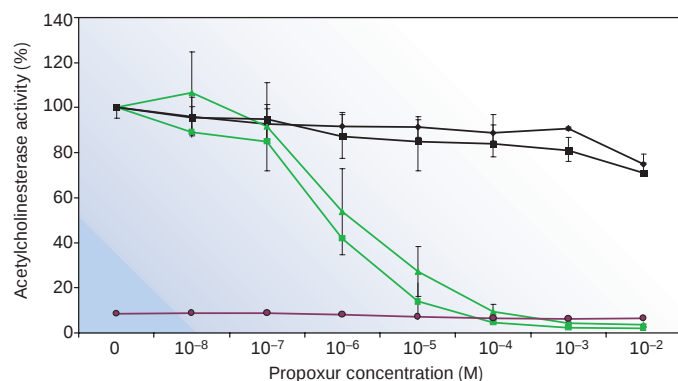


Figure 1 Residual acetylcholinesterase activity of susceptible (green squares) and resistant (black diamonds) mosquitoes assayed in homogenates and lysates from transfected S2 cells in the presence of increasing concentrations of the carbamate insecticide Propoxur. S2 cells were transfected with the recombinant pAc5.1/V5-His vector (Invitrogen) either alone (purple circles) or with expression of either sensitive acetylcholinesterase-1 (green triangles) or insensitive G119S-mutant enzyme (black squares). Residual enzyme activity was assayed after incubation with Propoxur for 15 min (ref. 5). Three independent experiments were carried out using different volumes of cell lysate.

carbamates¹. Here we show that this insensitivity results from a single amino-acid substitution in the enzyme, which we found in ten highly resistant strains of *C. pipiens* from tropical (Africa and Caribbean) and temperate (Europe) areas, as well as in one resistant African strain of *A. gambiae*. Our identification of this mutation may pave the way for designing new insecticides.

Acetylcholinesterase terminates synaptic transmission by hydrolysing the neurotransmitter acetylcholine; its inactivation by insecticides leads to paralysis and death. Mosquitoes, however, show widespread and strong resistance to this type of insecticide. They have two genes that encode different isoforms of acetylcholinesterase: *ace-1*, which has no homologue in the fruitfly *Drosophila melanogaster* and is closely linked to resistance in *C. pipiens*; and *ace-2*, a homologue of the unique *Drosophila ace* gene². The generally mild insensitivity of acetylcholinesterase-2 in *D. melanogaster* is due to the combined weak effect of several mutations³.

To identify mutations involved in resistance in mosquitoes, we determined the complete *ace-1* messenger RNA coding sequence of two *Culex pipiens* strains: one susceptible and one resistant (results not shown). *C. pipiens ace-1* encodes a putative 702-amino-acid protein, which is 81% identical to its *A. gambiae* homologue and 39% identical to *D. melanogaster* acetylcholinesterase-2. Complementary DNAs from the susceptible and resistant strains differ at 27 nucleotide positions, only one of which generates an amino-acid substitution: the GGC (glycine) codon at position 119, according to the nomenclature for *Torpedo* acetylcholinesterase¹, is replaced by an AGC (serine) codon in resistant mosquitoes (mutation G119S).

From three-dimensional modelling, we find that this mutated residue lies within the

active 'gorge' of the enzyme, close to the catalytic site and abutting the oxyanion hole (results not shown). To evaluate the biochemical effect of the mutation *in vitro*, we assayed the catalytic properties and insecticidal sensitivity of wild-type and mutant recombinant acetylcholinesterase-1 that was expressed in S2 *Drosophila* cells. The G119S mutant showed the same insensitivity to Propoxur insecticide as resistant-strain acetylcholinesterase-1 (Fig. 1). A single mutation in *ace-1* must therefore be responsible for the insensitivity of the enzyme.

To determine whether the G119S mutation is present in other *C. pipiens* strains with insensitive acetylcholinesterase, we sequenced exon 3 of *ace-1* in several resistant and susceptible strains derived either from the temperate or the tropical/subtropical form of the *C. pipiens* species complex (*C. p. pipiens* and *C. p. quinquefasciatus*, respectively). All of the resistant strains carried the G119S substitution, regardless of their origin. Moreover, although 23 nucleotides were polymorphic, a unique haplotype was found to be associated with the resistance within each subspecies (see supplementary information). This indicates that the same G119S mutation has occurred independently at least twice in *C. pipiens*, once in each subspecies.

We also investigated the recent emergence of insensitive acetylcholinesterase in the main African malaria vector *Anopheles gambiae*⁴, with the use of the *ace-1* genomic sequences of a resistant (YAO) and a susceptible (KISUMU) strain. The coding sequences differed at 18 nucleotide positions, two of them being non-synonymous. In the YAO strain, one mutation that resulted in the replacement of a valine residue by alanine in the amino-terminal region has no equivalent in *Torpedo* acetylcholinesterase and did not seem to affect the enzyme's catalytic properties (results not shown). The

other was the same G119S substitution as in *C. pipiens* (results not shown), indicating that this single point mutation has occurred independently at least three times in the *ace-1* gene: twice in the *C. pipiens* complex and once in *A. gambiae*.

The discovery of the *ace-1* mutation that is responsible for insecticide resistance in mosquitoes opens the way to new strategies for pest management. The development of new insecticides that can specifically inhibit the G119S mutant form of acetylcholinesterase-1 will be crucial in overcoming the spread of resistance.

Mylène Weill*, Georges Lutfalla†, Knud Mogensen‡, Fabrice Chandret, Arnaud Berthomieu*, Claire Berticat*, Nicole Pasteur*, Alexandre Philipps§, Philippe Fort§, Michel Raymond*

*Institut des Sciences de l'Évolution (UMR 5554), CC 065, Université Montpellier II, 34095 Montpellier, France

e-mail: weill@isem.univ-montp2.fr

†Défenses Antivirales et Tumorales (UMR 5124), and §Centre de Recherche en Biochimie des Macromolécules (UPR1086), CNRS, 34293 Montpellier, France

‡IRD/LIN, BP 64501, 34394 Montpellier, France

1. Toutant, J. P. *Progr. Neurobiol.* **32**, 423–446 (1989).
2. Weill, M. et al. *Proc. R. Soc. Lond. B* **269**, 2007–2016 (2002).
3. Mutero, A., Pralavorio, M., Bride, J. M. & Fournier, D. *Proc. Natl Acad. Sci. USA* **91**, 5922–5926 (1994).
4. N'Guessan, R. et al. *Med. Vet. Entomol.* **17**, 1–7 (2002).
5. Bourguet, D., Pasteur, N., Bisset, J. & Raymond, M. *Pest. Biochem. Physiol.* **55**, 122–128 (1996).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Nitrogen storage

UV-B radiation and soil microbial communities

Soil microorganisms regulate the supply of nitrogen to plants and so are important controllers of plant productivity and ecosystem carbon sequestration. Johnson *et al.*¹ report that exposure of a subarctic heath ecosystem to increased ultraviolet-B (UV-B) irradiation causes a drastic decline in the mass ratio of C:N in soil microorganisms, which would increase the amount of nitrogen stored in the microbial biomass and possibly alter the availability of nitrogen to plants. However, we argue that some of the authors' microbial C:N data are unrealistic, possibly because of an artefact of the technique used to measure microbial carbon and nitrogen concentrations. As a result, there is little reason to suppose that increased exposure of ecosystems to UV-B radiation will influence microbial nitrogen storage, plant nitrogen availability or rates of carbon sequestration.

Johnson *et al.* calculated an average microbial C:N ratio of 36.2 in control plots,

Table 1 Microbial biomass C:N ratios in soil from around the world

Ecosystem	Microbial carbon (mg per kg soil)	Microbial nitrogen (mg per kg soil)	C:N ratio
Temperate forest			
Organic layer	3,953	277	14.3
Mineral soil	807	93	8.7
Dry tropical forest	653	65	10.0
Tropical rainforest	986	100	9.9
Tropical savanna	342	35	9.8
Tussock grassland	1,288	161	8.0
Cool temperate arable	463	66	7.0
Warm temperate arable	331	47	7.0
Temperate managed grassland	1,011	170	5.9
Tropical arable	240	48	5.0

Data are based on a summary of values in ref. 3.

and a corresponding ratio of 8.2 in plots exposed to increased UV-B; this decline is the basis of their conclusion that UV-B increases microbial nitrogen storage. Although a C:N ratio of 8 is reasonable for soil microbial communities², and microbial C:N ratios of 20–29 have occasionally been reported, a ratio of 36 is substantially higher than any published value^{3,4} (Table 1).

Soil microbial C:N ratios determined in 119 temperate, tropical and boreal forest soils were found to be 8.2 ± 3.6 (mean $\pm$ s.d.) and 9.2 ± 3.3 in mineral soils and organic layers, respectively⁴. The values reported by Johnson *et al.* are therefore more than eight standard deviations higher than the mean for forest organic horizons, which themselves have high microbial C:N ratios relative to most soils (Table 1). Although few data are available for subarctic heath soils, mean microbial C:N ratios for a similar undisturbed heath soil in the vicinity of the plots used by Johnson *et al.* were found to vary seasonally and across several years between 4.6 and 9.9 (refs 5, 6). Plots that received large amounts of labile carbon had higher ratios, but even these averaged only 11.

The chloroform-fumigation extraction technique used by Johnson *et al.* to estimate microbial carbon and nitrogen is influenced by factors such as soil type, length of the chloroform-fumigation period, soil moisture content and the abundance of plant roots^{7,8}. By altering the soil moisture content during fumigation, or by extending the length of time that the soil is exposed to chloroform vapour, we were able to change the apparent microbial C:N ratio by a factor of two⁹.

We cannot explain why the microbial C:N ratios in the control plots reported by Johnson *et al.* are so high; they are out of the range of previously reported values for any soil microbial community, including those at another nearby subarctic heath site. This raises doubts about the authors' conclusion that UV-B irradiation alters nitrogen storage by soil microbes.

John M. Stark*, Stephen C. Hart†

*Department of Biology and the Ecology Center, Utah State University, Logan,

Utah 84322-5305, USA

e-mail: jstark@biology.usu.edu

†School of Forestry and the Merriam-Powell Center for Environmental Research, Northern Arizona University, Flagstaff, Arizona 86011-5018, USA

1. Johnson, D., Campbell, C. D., Lee, J. A., Callaghan, T. V. & Gwynn-Jones, D. *Nature* **416**, 82–83 (2002).
2. Paul, E. A. & Clark, F. E. *Soil Microbiology and Biochemistry* 2nd edn (Academic, San Diego, 1996).
3. Wardle, D. A. *Biol. Rev.* **67**, 321–358 (1992).
4. Bauhus, J. & Khanna, P. K. in *Going Underground: Ecological Studies in Forest Soils* (eds Rastin, N. & Bauhus, J.) 78–110 (Research Signpost, Trivandrum, India, 1999).
5. Jonasson, S., Michelsen, A., Schmidt, I. K., Nielson, E. V. & Callaghan, T. V. *Oecologia* **106**, 507–515 (1996).
6. Michelsen, A. *et al.* *New Phytol.* **143**, 523–538 (1999).
7. Haubensak, K. A., Hart, S. C. & Stark, J. M. *Soil Biol. Biochem.* **34**, 1549–1562 (2002).
8. Mueller, T., Joergensen, R. G. & Meyer, B. *Soil Biol. Biochem.* **24**, 179–181 (1992).

Johnson *et al.* reply — Stark and Hart question our finding¹ that increased UV-B irradiation affects the biomass C:N ratio of Arctic soil microbial communities. Our data for microbial carbon biomass (C_{mic} ; 1.0–3.1 mg per g soil) and nitrogen (N_{mic} ; 0.1–0.3 mg per g soil) fall within the range presented by Stark and Hart (0.2–3.9 and 0.04–0.28 mg per g soil for C_{mic} and N_{mic} , respectively)², but the C:N ratios in the control plots are greater than those for both similar^{3,4} and contrasting soil types⁵. However, we do not agree that this invalidates our conclusions.

Stark and Hart suggest that the high C:N ratios could have occurred as a result of artefacts in the experimental conditions or differences in basic soil properties, such as moisture or carbon content, during fumigation. The finding that basic soil properties can affect fumigation⁶ suggests that there is limited value in comparing C_{mic} and N_{mic} values from different studies. We agree that this could have important consequences for the results if samples within a study were to vary significantly in these properties. However, we subsequently found no correlation between the microbial C:N ratio and moisture content ($P = 0.218$) or carbon content ($P = 0.300$) of the original samples.

For statistical analysis, we calculated separate C:N ratios for each pair of C_{mic} and N_{mic} values and determined the means

of these ratios. However, the C:N ratios presented by Stark and Hart are calculated from a mean C_{mic} value and a mean N_{mic} value. Calculation of our C:N ratios using their method gives ratios of 24.1 and 8.9 for ambient and increased UV-B, respectively. These values are much lower for the ambient UV-B treatment, are within the range of published values^{2–4,6,7}, and still follow the same pattern as those that we originally reported.

Furthermore, we discussed the C:N ratios together with complementary and independent data, which showed significant changes in utilization of carbon compounds by soil bacteria in response to increased UV-B irradiation. These combined results probably reflect alterations in the composition of the microbial community. Our conclusion that increased UV-B irradiation may affect microbial nitrogen storage was based on the increase in N_{mic} .

Stark and Hart's observations highlight the need for more research on the structure and functioning of microbial communities and the methods used to study them. This may be particularly relevant for subarctic heaths, for which there are few comparable data. Previous work^{3,4} from sites close to ours reported C_{mic} and N_{mic} values (5.6–12.8 and 0.7–1.3 mg per g soil, respectively) that were highly variable and also greater than those listed by Stark and Hart. Together with our values, these indicate that the microbial biomass and C:N ratio in these extremely organic soils is highly variable and site-specific.

We therefore contend that our data, which were obtained under identical experimental conditions, provide evidence that increased UV-B irradiation can affect both the community structure and the amount of nitrogen held within the microbial biomass.

David Johnson*, Colin D. Campbell†, John A. Lee*, Terry V. Callaghan*‡, Dylan Gwynn-Jones§

*Department of Animal and Plant Science, University of Sheffield, Western Bank, Sheffield S10 2TN, UK
e-mail: d.johnson@shef.ac.uk

†The Macaulay Land Use Research Institute, Craigiebuckler, Aberdeen AB15 8QH, UK

‡Abisko Scientific Research Station, 981-07 Abisko, Sweden

§Institute of Biological Sciences, University of Wales, Aberystwyth SY23 3DA, UK

1. Johnson, D., Campbell, C. D., Lee, J. A., Callaghan, T. V. & Gwynn-Jones, D. *Nature* **416**, 82–83 (2002).
2. Wardle, D. A. *Biol. Rev.* **67**, 321–358 (1999).
3. Michelson, A. *et al.* *New Phytol.* **143**, 523–538 (1999).
4. Jonasson, S., Michelsen, A., Schmidt, I. K., Nielson, E. V. & Callaghan, T. V. *Oecologia* **106**, 507–515 (1996).
5. Haubensak, K. A., Hart, S. C. & Stark, J. M. *Soil Biol. Biochem.* **34**, 1549–1562 (2002).
6. Shibahara, F. & Inubushi, K. *Soil Sci. Plant Nutr.* **41**, 681–689 (1995).
7. Inubushi, K., Brookes, P. C. & Jenkinson, D. S. *Soil Biol. Biochem.* **23**, 737–741 (1991).

The evolutionary origin of complex features

Richard E. Lenski*, Charles Ofria†, Robert T. Pennock‡ & Christoph Adami§

* Department of Microbiology & Molecular Genetics, † Department of Computer Science & Engineering, and ‡ Lyman Briggs School & Department of Philosophy, Michigan State University, East Lansing, Michigan 48824, USA

§ Digital Life Laboratory, California Institute of Technology, Pasadena, California 91125, USA

A long-standing challenge to evolutionary theory has been whether it can explain the origin of complex organismal features. We examined this issue using digital organisms—computer programs that self-replicate, mutate, compete and evolve. Populations of digital organisms often evolved the ability to perform complex logic functions requiring the coordinated execution of many genomic instructions. Complex functions evolved by building on simpler functions that had evolved earlier, provided that these were also selectively favoured. However, no particular intermediate stage was essential for evolving complex functions. The first genotypes able to perform complex functions differed from their non-performing parents by only one or two mutations, but differed from the ancestor by many mutations that were also crucial to the new functions. In some cases, mutations that were deleterious when they appeared served as stepping-stones in the evolution of complex features. These findings show how complex functions can originate by random mutation and natural selection.

Charles Darwin's theory of evolution¹, including its intertwined hypotheses of descent with modification and adaptation by natural selection, is widely regarded as one of the greatest scientific achievements of all time. From the outset, Darwin realized that “organs of extreme perfection and complication”, such as the eye, posed a difficulty for his theory. Such features are much too complex to appear *de novo*, and he reasoned that they must evolve by incremental transitions through many intermediate states, sometimes undergoing changes in function. There now exists substantial evidence concerning the evolution of complex features that supports Darwin's general model^{2–16}. Nonetheless, it is difficult to provide a complete account of the origin of any complex feature owing to the extinction of intermediate forms, imperfection of the fossil record, and incomplete knowledge of the genetic and developmental mechanisms that produce such features.

To examine the evolutionary origin of a complex feature in much greater detail than has previously been possible, we have performed experiments with digital organisms—computer programs that self-replicate, mutate and compete^{17–26}. As Daniel Dennett²⁷ has emphasized, “...evolution will occur whenever and wherever three conditions are met: replication, variation (mutation), and differential fitness (competition)”. By using this tractable system, we aim to shed light on principles relevant to any evolving system. Digital organisms are also of interest as computer scientists and engineers explore ways to apply evolutionary principles to program design, engineering and robotics^{28–31}. The next section describes our experimental system, including the logic functions that digital organisms can use to obtain energy. There follows a case study of the genomic and phenotypic changes in a population that evolved an especially complex function, and then a functional-genomic analysis of the first genotype able to perform that function. Next we compare the trajectories and solutions of many populations that independently evolved this same function, and we conclude by examining the influence of different selective environments on the propensity of digital organisms to evolve the function.

Experimental system

Avida is a software platform for research on digital organisms¹⁹. To convey what digital organisms are, it is helpful to compare them with computer viruses. Digital organisms are self-replicating computer programs, as are viruses. However, computer viruses require

direct intervention to mutate and evolve, whereas digital organisms exist in a computational environment where copying is imperfect such that they mutate randomly and evolve spontaneously. Also, digital organisms compete for energy and, depending on the environment, can obtain energy by performing logic functions^{19,25}.

In Avida, a genome is a circular sequence of instructions (Fig. 1) that are executed sequentially except when the execution of one instruction causes the instruction pointer to jump to another position. Jumps are encoded using a template-based system, which is more robust to mutation than is numerical addressing^{17,24}. Each item in a sequence is one of 26 possible instructions (Supplementary Information). Reproduction is asexual and occurs by binary fission. Experiments began with an ancestor that could replicate but could not perform any logic functions. All organisms

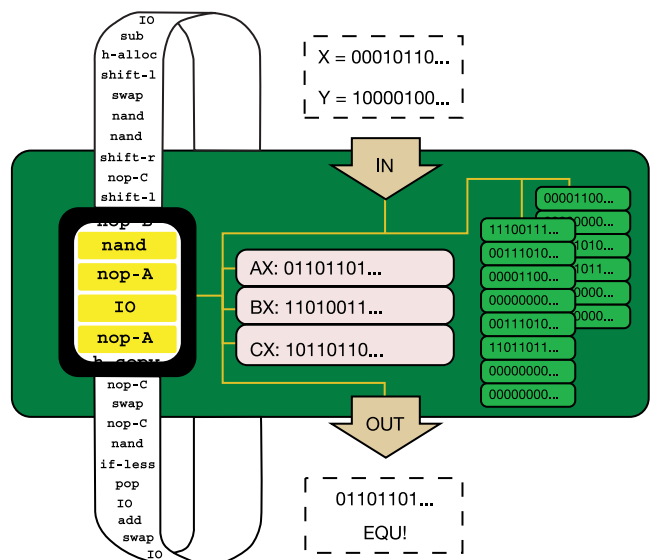


Figure 1 Digital organism in Avida. Each individual organism has a circular genome and a virtual CPU with two stacks and three registers that hold 32-bit strings. Execution of the genomic program generates a computational metabolism, whereby numerical substrates can be input from the environment, processed in stacks and registers, and resulting products output to the environment. See main text for details.

were identical and obtained equal energy to execute their genomic programs, including the copy commands by which a genome replicates itself one instruction at a time. Copying is subject to errors, including point mutations, insertions and deletions. Each mutation alters the genome and may change an organism's phenotype, including its replication efficiency, computational metabolism and robustness. Thus, genotypes vary in their expected reproductive success. As in nature, selection in Avida depends on the phenotypic effects of a mutation in its genetic context and in relation to the organism's environment; the researcher does not specify a distribution of selection coefficients. Most mutations in Avida are deleterious or neutral, but a small fraction increases fitness²⁰.

Digital organisms compete for the energy needed to execute instructions. Energy occurs as discrete quanta called 'single-instruction processing' units, or SIPs. Each SIP suffices to execute one instruction. By executing instructions, a digital organism can express phenotypes that enable it to obtain more energy and copy its genome. In Avida, organisms can acquire energy by two mechanisms. First, each organism receives SIPs in proportion to its genome length. Second, an organism can obtain further SIPs by performing one- and two-input logic operations on 32-bit strings (Supplementary Information). Only one of the 26 instructions in the genetic code, `nand` ('not and'), is itself a logic operator; and `nand` must be executed in coordination with `IO` ('input-output') instructions to perform the NAND function. All other logic functions can be constructed using one or more `nand` instructions within an integrated framework of other instructions.

Two logic functions, NAND and EQU ('equals'), are shown in Fig. 1. The execution of the highlighted `nand` instruction, when immediately followed by the modifying `nop-A` ('no operation') instruction, causes the bit-strings in the BX and CX registers to be combined according to the `nand` operator and the result to be written to the AX register. The `nand` operator returns 0 ('false') when both inputs are 1 ('true'); it returns 1 if one or both inputs are 0. The subsequent `IO` instruction and its modifying `nop-A` cause the string in the AX register to be output. If this output matches perfectly the correct answer for a function that is rewarded, then the organism's rate of energy acquisition, and hence the execution of its genomic program, is accelerated by the factor shown in Table 1. For example, if the organism in Fig. 1 had previously input the strings labelled X and Y, and if it now output the string in the AX register (generated by the preceding `nand` and `nop-A` instructions), then the organism would receive the reward for performing EQU. The reward is obtained because the output string has a 1 at every position where X and Y are equal (both 0 or both 1), and it has a 0 at each position where X and Y differ. The organism would obtain no reward if any of the 32 bits in the string were incorrect, or if it had already been rewarded in its lifetime for performing EQU.

Our study focuses on the origin of the EQU function. EQU garners the largest reward in our experiments because its performance is more complex than any other one- or two-input logic

function, given the available genomic instructions. An exhaustive search shows that the minimum number of `nand` operations to perform EQU is five, which is greater than for any other one- or two-input logic function. Using the 26 available instructions, we wrote a program of length 19 that performs EQU but does not replicate (Supplementary Information). This program seems, but has not been proven, to be the shortest one to perform EQU.

Case-study population

The ancestor could replicate but could not perform any logic function. However, an organism that evolved one or more of nine logic functions would obtain further energy. The benefit increased exponentially with the approximate difficulty of each function (Table 1). Functions could be performed in any order during an individual's life, but no extra energy was obtained by repeatedly performing the same function. No single mutation in the ancestor can produce even the simplest of these functions. Instead, several mutant instructions must appear in the same lineage, and such that they are coordinately executed, to perform even a simple function. Nonetheless, this population (and many others) evolved the capacity to perform EQU, the most complex of these functions.

Figure 2 shows the trajectory of the population's divergence from the ancestor. The vertical axis, phylogenetic depth, is the cumulative number of generations in which an individual differed from its parent by one or more mutations. The horizontal axis is time in computational updates (see Methods). The colour scale shows the abundance of genotypes at a given depth. The blue line shows the exact line of descent leading to the genotype that was most abundant at the end of the experiment. This final dominant type was 344 steps removed from its ancestor. That is, an offspring differed genetically from its parent in 344 of the many thousands of generations leading to this final type. The final dominant has a genome 83 instructions long (the ancestral length was 50), and it performs all nine logic functions that provide energy.

Figure 3 shows the trajectories for two fitness components, replication efficiency and computational merit, for all 344 geno-

Table 1 Rewards for performing nine one- and two-input logic functions

Function name	Logic operation	Computational merit
NOT	$\sim A$; $\sim B$	2
NAND	$\sim(A \text{ and } B)$	2
AND	$A \text{ and } B$	4
OR_N	$(A \text{ or } \sim B)$; $(\sim A \text{ or } B)$	4
OR	$A \text{ or } B$	8
AND_N	$(A \text{ and } \sim B)$; $(\sim A \text{ and } B)$	8
NOR	$\sim A \text{ and } \sim B$	16
XOR	$(A \text{ and } \sim B) \text{ or } (\sim A \text{ and } B)$	16
EQU	$(A \text{ and } B) \text{ or } (\sim A \text{ and } \sim B)$	32

The symbol ' $\sim$ ' denotes negation. The reward for computational merit increases with 2^n , where n is the minimum number of `nand` operations needed to perform the listed function. Symmetrical operations, shown separated by a semi-colon, are treated as the same function. No added benefit is obtained for performing any function multiple times. These functions include all one- and two-input logic operations except ECHO, which requires no `nand` operations and was not rewarded.

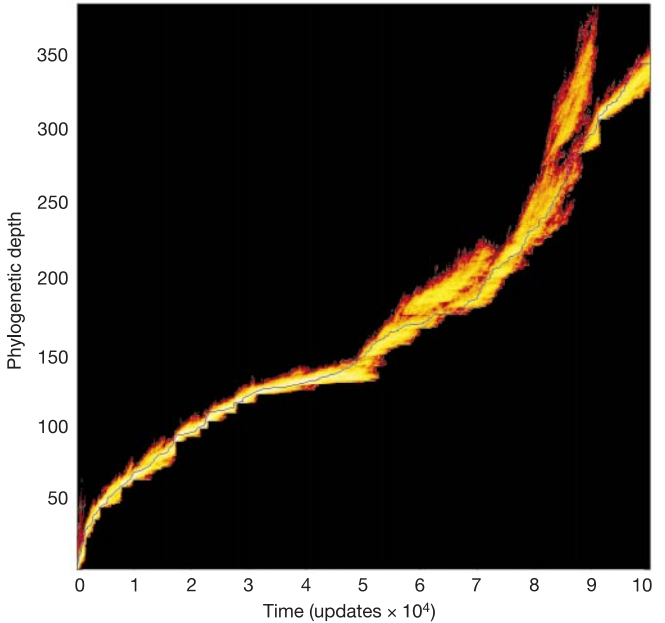


Figure 2 Phylogenetic depth versus time in the case-study population. Phylogenetic depth is the cumulative number of generations in which an organism's genotype differs from its parent. The exact line of descent leading to the most abundant final genotype is shown as the blue line. Colours indicate the relative abundance of genotypes at any depth, yellow being more abundant than red.

types along the line of descent. Replication efficiency is the ratio of an organism's genome length to the SIPs used during its life cycle. Computational merit is the total reward obtained over an organism's lifetime for performing logic functions. Each genotype's expected replication rate, or fitness, equals the product of these quantities. We also identified every mutation that fell along the line of descent in this population, and characterized the phenotypic changes associated with each step (Supplementary Information). The EQU function first appeared at step 111 (update 27,450). There were 103 single mutations, six double mutations, and two triple mutations among these steps. Forty-five of the steps increased overall fitness, 48 were neutral and 18 were deleterious relative to the immediate parent. The large proportion of beneficial mutations along the line of descent is not surprising, because this lineage represents the eventual winners that were assembled by natural selection. Thirteen of the 45 beneficial steps gave rise to the expression of logic functions not expressed by the immediate parent, including six steps in which there were trade-offs of simpler for more complex functions.

The presence of deleterious mutations along the line of descent is more surprising. Fifteen of the 18 deleterious mutations reduced fitness by $<3\%$ relative to the parent, and might have hitchhiked with beneficial mutations that arose soon after in the same genetic background. However, two mutations reduced fitness by $>50\%$. One was a point mutation that disrupted replication efficiency. Its harmful effect was eliminated by the next mutation in the line of descent, which occurred at a distant site in the genome. The other very deleterious step was a point mutation, at depth 110, that knocked out NAND, one of the simplest logic functions. Only two individuals had this maladapted genotype, yet their descendants emerged as eventual winners. In fact, in the very next step, this genotype produced the mutation that gave rise to EQU. Was that deleterious mutation extremely lucky to hitchhike with such a

beneficial mutation? Or was the deleterious mutation a prerequisite for producing the EQU function within that genome context? To distinguish between these hypotheses, we reversed this one-step-prior mutation in the genotype that first expressed EQU. This reversal eliminated the EQU function. Therefore, a mutation that was highly deleterious when it appeared was highly beneficial in combination with a subsequent mutation. The evolution of a complex feature, such as EQU, is not always an inexorably upward climb toward a fitness peak, but instead may involve sideways and even backward steps, some of which are important.

After the origin of EQU, another 233 steps occurred along the line of descent leading to the final dominant genotype. Of these, 62 were beneficial, 132 neutral and 39 deleterious. All nine logic functions were performed from genotype 306 onwards.

Functional genomics

To determine how many instructions were required to perform EQU at its origin in the case-study population, we systematically replaced each existing instruction with a null instruction in the first genotype able to perform EQU. We scored each null mutant for which functions were lost and which remained, and the data from all the mutants were combined to produce an array showing the relationship between genome sequence and phenotypic properties (Fig. 4). The genome of the first EQU-performing organism had 60 instructions; eliminating any of 35 of them destroyed that function. Although the mutation of only one instruction produced this innovation when it originated, the EQU function evidently depends on many interacting components.

The many instructions required to perform EQU, the previous evolution of simpler functions, and the repeated trade-offs between simpler and more complex functions all suggest functional interconnections between the genetic networks encoding them. We used the same array to ask how many instructions that were required to perform EQU were also required for other functions. Besides EQU, this genotype performed five of the eight simpler logic functions; AND was lost as a side-effect of the EQU-producing mutation, and NAND had been eliminated by the one-step-prior mutation. Of the 35 instructions required for EQU, 22 were needed for simpler functions. Three instructions required for EQU were also essential for replication; these were conserved from the ancestral sequence, as were five others. However, 27 of the 35 instructions required to perform EQU were evolutionarily derived, and all but one of them had appeared in the line of descent before this function was ever performed. Thus, although more than two dozen mutations were used to build EQU, undoing any one of them destroyed this function. This asymmetry might suggest that EQU is fragile and cannot last. However, the EQU function lasted throughout the experiment because, once evolved, it was so valuable that defective mutants were eliminated by selection.

Variations on a theme

The case-study population was one of 50 that evolved under identical conditions, 23 of which acquired EQU. The phylogenetic depth at which EQU first appeared ranged from 51 to 721 steps. In principle, 16 mutations, coupled with three instructions already present in the ancestor, could have produced an EQU-performing organism. The actual paths were much longer and highly variable, indicating the circuitousness and unpredictability of evolution leading to this complex feature. We identified the pivotal genotypes that first performed EQU in each population, and the pivotal mutations that distinguished them from their parents. We use 'pivotal' to mark these milestones, not to imply that earlier genotypes and mutations were unimportant for the origin of EQU. A single mutation distinguished the pivotal genotype from its parent in 19 populations, whereas four involved double mutations. The pivotal mutations included point mutations, insertions, a small duplication and even deletions. Pivotal point

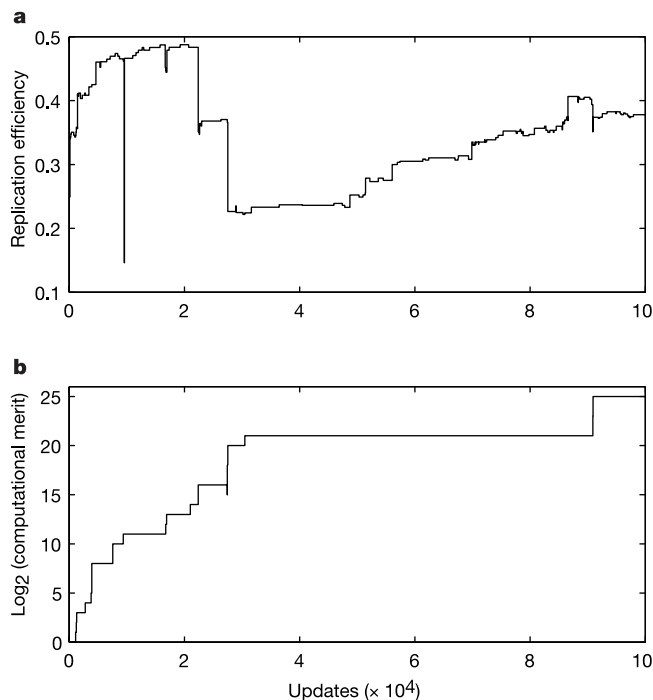


Figure 3 Trajectories for two fitness components, showing each genotype in the line of descent for the case-study population. **a**, Replication efficiency, which is the ratio of genome length to energy used in an organism's lifetime. **b**, Computational merit, shown $\log_2$ -transformed, which is the product of all the rewards obtained by an organism for logic functions performed during its lifetime.

Instruction	Repl	NOT	NAND	AND	OR_N	OR	AND_N	NOR	XOR	EQU
1 r h-alloc										
2 m dec										
3 z set-flow										
4 a nop-A										
5 v mov-head										
6 c nop-C										
7 g push										
8 m dec										
9 c nop-C										
10 l swap										
11 q IO										
12 q IO										
13 p nand										
14 t h-copy										
15 q IO										
16 p nand										
17 q IO										
18 c nop-C										
19 p nand										
20 c nop-C										
21 t h-copy										
22 l inc										
23 e if-less										
24 t h-copy										
25 n add										
26 c nop-C										
27 o sub										
28 g push										
29 c nop-C										
30 b nop-B										
31 e if-less										
32 a nop-A										
33 m dec										
34 q IO										
35 d if-n-equ										
36 t h-copy										
37 q IO										
38 c nop-C										
39 p nand										
40 t h-copy										
41 l swap										
42 p nand										
43 q IO										
44 f pop										
45 p nand										
46 g push										
47 q IO										
48 x get-head										
49 u h-search										
50 t h-copy										
51 y if-label										
52 c nop-C										
53 u h-search										
54 a nop-A										
55 s h-divide										
56 t h-copy										
57 t h-copy										
58 t h-copy										
59 v mov-head										
60 a nop-A										
State changes	8	3	3	7	7	19	12	13	0	35

mutations used 10 of the 26 possible instructions; and several pivotal insertions and deletions were unique. Therefore, many different mutations produced the EQU function in the right genomic context.

Before the pivotal mutations, each line of descent had already experienced 50 or more steps, including beneficial, neutral and deleterious mutations. Some deleterious mutations had small effects and hitchhiked with beneficial mutations; others reverted or were compensated by mutations elsewhere. But in the case study, we showed that one deleterious mutation had genetically predisposed (through an epistatic interaction) the subsequent origin of EQU. To investigate this possibility further, we examined all of the mutations one step before the pivotal mutations. Five of the 23 one-step-prior mutations were deleterious in the backgrounds in which they occurred (including the case study). When these five were reverted in the pivotal genotypes, the EQU function was eliminated in three cases. Thus, three genotypes that first performed this complex function depended on a one-step-prior mutation that was deleterious when it appeared.

The 23 pivotal genomes ranged in length from 49 to 356 instructions; the ancestor had 50 instructions and so there was a strong tendency for increased length to precede the origin of EQU. The parents of the pivotal genotypes performed between four and all eight of the simpler logic functions, with a median of seven. Therefore, at least several simpler functions evolved before EQU in every population. However, no particular function was essential for the subsequent evolution of EQU, because each one was absent from at least one of the 23 parent genotypes. All the pivotal mutations were beneficial owing to the extra energy obtained by the EQU function. However, net gains also depended on pleiotropic side-effects on other traits. Twenty of the 23 pivotal mutations caused losses of one or more logic functions the parent performed, and 13 reduced replication efficiency. Eight yielded another function, and three improved replication efficiency. On balance, 20 of the 23 pivotal mutations would have been deleterious if EQU had not been rewarded, whereas three would still have been beneficial. Thus, beneficial mutations that produced this new function usually, but not always, engendered trade-offs in other aspects of performance.

We ran the functional-genomic analyses on all 23 pivotal genotypes. The number of instructions required for EQU ranged from 17 to 43, with a median of 28 instructions. Notice that one evolved type apparently needed only 17 instructions to perform EQU, whereas our shortest hand-written program used 19 instructions. Further examination showed that this unexpectedly low value occurred because the evolved genome was partially redundant, such that one-at-a-time null mutations did not reveal the full extent of its computational network. These analyses therefore provide a minimum estimate of the number of instructions that a genotype uses to perform a function.

Different environments

We carried out experiments to examine the effects of different selective environments on the propensity to evolve the EQU function, with all other conditions held constant. Ten further populations evolved under each of 36 possible regimes in which one or

Figure 4 Functional-genomic array for the first organism to perform EQU in the case-study population. Its genome sequence is shown to the left; the instruction highlighted in yellow is the pivotal mutation that yielded EQU but simultaneously eliminated AND. Top labels denote replication (Repl.) and logic functions; associated colours show whether this organism can (green) or cannot (red) perform the function. The fill in each interior cell shows the effect on the function of replacing the instruction with a null instruction. Red, null mutation destroys existing function; blank, null mutation has no qualitative effect; green, null mutation produces new function. The number of state changes for each function is shown at the bottom.

two simpler functions were not rewarded. In all environments, at least one population evolved EQU. Evidently, neither any particular simpler function nor any pairwise combination of functions was required to evolve this complex feature. In these 36 environments, the overall fraction of populations that evolved EQU was 124 of 360 (34%), only slightly less than in the 'reward-all' environment ($P = 0.0764$, one-tailed Fisher's exact test).

At the other extreme, 50 populations evolved in an environment where only EQU was rewarded, and no simpler function yielded energy. We expected that EQU would evolve much less often because selection would not preserve the simpler functions that provide foundations to build more complex features. Indeed, none of these populations evolved EQU, a highly significant difference from the fraction that did so in the reward-all environment ($P \approx 4.3 \times 10^{-9}$, Fisher's exact test). However, these populations tested more genotypes, on average, than did those in the reward-all environment (2.15×10^7 versus 1.22×10^7 ; $P < 0.0001$, Mann-Whitney test), because they tended to have smaller genomes, faster generations, and thus turn over more quickly. However, all populations explored only a tiny fraction of the total genotypic space. Given the ancestral genome of length 50 and 26 possible instructions at each site, there are $\sim 5.6 \times 10^{70}$ genotypes; and even this number underestimates the genotypic space because length evolves.

Discussion

By using digital organisms, we traced the exact genealogy, without any 'missing links', from an ancestor that could replicate only to descendants able to perform multiple logic functions requiring the coordinated execution of many genomic instructions. The most complex function, EQU, evolved only when several simpler functions were also useful. Some simpler functions were accessible from the ancestor by relatively few mutations, and these served as a foundation on which more complex features were built. The foundational role of simpler functions in the origin of more complex ones was evident in the overlap of the genetic networks underlying their expression, and the frequent loss of simpler functions as side-effects of mutations yielding more complex functions. Our experiments demonstrate the validity of the hypothesis, first articulated by Darwin¹ and supported today by comparative and experimental evidence^{2–16}, that complex features generally evolve by modifying existing structures and functions. Some readers might suggest that we 'stacked the deck' by studying the evolution of a complex feature that could be built on simpler functions that were also useful. However, that is precisely what evolutionary theory requires, and indeed, our experiments showed that the complex feature never evolved when simpler functions were not rewarded. Our experiments also show that many different genomic solutions produce the same complex function. Following any particular path is extremely unlikely, but the complex function evolved with a high probability, implying a very large number of potential paths³². Although the complex feature first appeared as the immediate result of only one or two mutations, its function invariably depended on many instructions that had previously evolved to perform other functions, such that their removal would eliminate the new feature.

Of course, digital organisms differ from organic life in their genetic constitution, metabolic activities and physical environments. However, digital organisms undergo the same processes of reproduction, mutation, inheritance and competition that allow evolution and adaptation by natural selection in organic forms. Other similarities are pleiotropy (one mutation affects multiple traits) and epistasis (multiple mutations interact to determine the same trait), which emerge in both organic and digital life from the nonlinear processes by which genomes encode and produce phenotypic features. As a consequence, selection acts on organisms rather than directly on their genes. The organisms in our study were asexual. Sex might accelerate the evolution of complex features by combining functions that independently evolved in different

lineages. On the other hand, asexuality permits beneficial combinations of mutations to spread even when they are individually deleterious, as sometimes occurred in our experiments. The consequences of asexual and sexual reproduction for the evolution of complex features deserve further research. In closing, digital organisms provide opportunities to address important issues in evolutionary biology. They are particularly well suited to problems that are difficult to study with organic forms owing to incomplete information, insufficient time and the impracticality of experiments. □

Methods

Avida software

Avida uses a time-slicing algorithm¹⁹ to ensure that all organisms execute instructions in an effectively parallel manner. Experiments ran using Avida version 1.6 on the Linux operating system on a Beowulf cluster of 64 Pentium III processors. The software and configuration files for our experiments can be obtained free from our website (<http://myxo.css.msu.edu/papers/nature2003>). More information about Avida, and further data from our experiments, can also be found there.

Experimental conditions

Every population started with 3,600 identical copies of an ancestral genotype that could replicate but could not perform any logic functions. Each replicate population that evolved in the same environment was seeded with a different random number. The hand-written ancestral genome was 50 instructions long, of which 15 were required for efficient self-replication; the other 35 were tandem copies of a single no-operation instruction (nop-c) that performed no function when executed. Copy errors caused point mutations, in which an existing instruction was replaced by any other (all with equal probability), at a rate of 0.0025 errors per instruction copied. Single-instruction deletions and insertions also occurred, each with a probability of 0.05 per genome copied. Hence, in the ancestral genome of length 50, 0.225 mutations are expected, on average, per replication. Various organisms from nature have genomic mutation rates higher or lower than this value³³. Mutations in Avida also occasionally cause the asymmetrical division of a copied genome, leading to the deletion or duplication of multiple instructions. Each digital organism obtained 'energy' in the form of SIPs at a relative rate (standardized by the total demand of all organisms in the population) equal to the product of its genome length and computational merit, where the latter is the product of rewards for logic functions performed. The exponential reward structure shown in Table 1 was used in the reward-all environment, whereas some functions obtained no reward under other regimes. An organism's expected reproductive rate, or fitness, equals its rate of energy acquisition divided by the amount of energy needed to reproduce. Fitness can also be decomposed into replication efficiency (ratio of genome length to energy required for replication) and computational merit. Each population evolved for 100,000 updates, an arbitrary time unit equal to the execution of 30 instructions, on average, per organism. The ancestor used 189 SIPs to produce an offspring, so each run lasted for 15,873 ancestral generations. Populations existed on a lattice with a capacity of 3,600 individuals. When an organism copied its genome and divided, the resulting offspring was randomly placed in one of the eight adjacent cells or in the parent's cell. Each birth caused the death of the individual that was replaced, thus maintaining a constant population size.

Received 19 September 2002; accepted 13 March 2003; doi:10.1038/nature01568.

1. Darwin, C. *On the Origin of Species by Means of Natural Selection* (Murray, London, 1859).
2. Jacob, F. Evolution and tinkering. *Science* **196**, 1161–1166 (1977).
3. Salvini-Plawen, L. V. & Mayr, E. On the evolution of photoreceptors and eyes. *Evol. Biol.* **10**, 207–263 (1977).
4. Mortlock, R. R. (ed.) *Microorganisms as Model Systems for Studying Evolution* (Plenum, New York, 1984).
5. Dawkins, R. *The Blind Watchmaker* (Norton, New York, 1986).
6. Goldsmith, T. H. Optimization, constraint, and history in the evolution of eyes. *Q. Rev. Biol.* **65**, 281–322 (1990).
7. Piatigorsky, J. & Wistow, G. The recruitment of crystallins: new functions precede gene duplication. *Science* **252**, 1078–1079 (1991).
8. Land, M. F. & Fernald, R. D. The evolution of eyes. *Annu. Rev. Neurosci.* **15**, 1–29 (1992).
9. Nilsson, D.-E. & Pelger, S. A pessimistic estimate of the time required for an eye to evolve. *Proc. R. Soc. Lond. B* **256**, 53–58 (1994).
10. Meléndez-Hevia, E., Waddell, T. G. & Cascante, M. The puzzle of the Krebs citric acid cycle: assembling the pieces of chemically feasible reactions, and opportunism in the design of metabolic pathways during evolution. *J. Mol. Evol.* **43**, 293–303 (1996).
11. Chen, L., DeVries, A. L. & Cheng, C.-H. C. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl Acad. Sci. USA* **94**, 3811–3816 (1997).
12. Dean, A. M. & Golding, G. B. Protein engineering reveals ancient adaptive replacements in isocitrate dehydrogenase. *Proc. Natl Acad. Sci. USA* **94**, 3104–3109 (1997).
13. Newcomb, R. D. et al. A single amino acid substitution converts a carboxylesterase to an organophosphorus hydrolase and confers insecticide resistance on a blowfly. *Proc. Natl Acad. Sci. USA* **94**, 7464–7468 (1997).
14. Miller, K. R. *Finding Darwin's God* (Cliff Street, New York, 1999).
15. Yokoyama, S. & Radlwimmer, F. B. The molecular genetics and evolution of red and green colour vision in vertebrates. *Genetics* **158**, 1697–1710 (2001).
16. Wilkins, A. S. *The Evolution of Developmental Pathways* (Sinauer, Sunderland, Massachusetts, 2002).

17. Ray, T. S. in *Artificial Life II* (eds Langton, C. G., Taylor, C., Farmer, J. D. & Rasmussen, S.) 371–408 (Addison–Wesley, Redwood City, California, 1991).
18. Bedau, M. A., Snyder, E., Brown, C. T. & Packard, N. H. in *Proc. 4th Eur. Conf. Artif. Life* (eds Husband, P. & Harvey, I.) 125–134 (MIT Press, Cambridge, Massachusetts, 1997).
19. Adami, C. *Introduction to Artificial Life* (Springer, New York, 1998).
20. Lenski, R. E., Ofria, C., Collier, T. C. & Adami, C. Genome complexity, robustness and genetic interactions in digital organisms. *Nature* **400**, 661–664 (1999).
21. Wagenaar, D. & Adami, C. in *Artificial Life VII* (eds Bedau, M. A., McCaskill, J. S., Packard, N. H. & Rasmussen, S.) 216–220 (MIT Press, Cambridge, Massachusetts, 2000).
22. Wilke, C. O., Wang, J. L., Ofria, C., Lenski, R. E. & Adami, C. Evolution of digital organisms at high mutation rate leads to survival of the flattest. *Nature* **412**, 331–333 (2001).
23. Yedid, G. & Bell, G. Microevolution in an electronic microcosm. *Am. Nat.* **157**, 465–487 (2001).
24. Ofria, C., Adami, C. & Collier, T. C. Design of evolvable computer languages. *IEEE Trans. Evol. Comput.* **6**, 420–424 (2002).
25. Wilke, C. O. & Adami, C. The biology of digital organisms. *Trends Ecol. Evol.* **17**, 528–532 (2002).
26. Yedid, G. & Bell, G. Macroevolution simulated with autonomously replicating computer programs. *Nature* **420**, 810–812 (2002).
27. Dennett, D. in *Encyclopedia of Evolution* (ed. Pagel, M.) E83–E92 (Oxford Univ. Press, New York, 2002).
28. Holland, J. H. *Adaptation in Natural and Artificial Systems* (MIT Press, Cambridge, Massachusetts, 1992).
29. Koza, J. R. *Genetic Programming* (MIT Press, Cambridge, Massachusetts, 1992).
30. Lipson, H. & Pollack, J. B. Automatic design and manufacture of robotic life forms. *Nature* **406**, 974–978 (2000).
31. Foster, J. A. Evolutionary computation. *Nature Rev. Genet.* **2**, 428–436 (2001).
32. Dennett, D. C. *Darwin's Dangerous Idea* (Simon & Schuster, New York, 1995).
33. Drake, J. W., Charlesworth, B., Charlesworth, D. & Crow, J. F. Rates of spontaneous mutation. *Genetics* **148**, 1667–1686 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Bennett, J. Bull, J. Coyne, D. Lenski, M. Lenski and E. Zuckerkandl for comments. The authors' work is supported by the US National Science Foundation Biocomplexity Program and by the MSU Foundation. Part of this work was carried out at the Jet Propulsion Laboratory under contract with the US National Aeronautics and Space Administration.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.E.L. (lenski@msu.edu).

The structure of DNA in the nucleosome core

Timothy J. Richmond & Curt A. Davey

ETH Zürich, Institut für Molekularbiologie und Biophysik, ETH-Hönggerberg, CH-8093 Zürich
The authors contributed equally to this work

The 1.9-Å-resolution crystal structure of the nucleosome core particle containing 147 DNA base pairs reveals the conformation of nucleosomal DNA with unprecedented accuracy. The DNA structure is remarkably different from that in oligonucleotides and non-histone protein–DNA complexes. The DNA base-pair-step geometry has, overall, twice the curvature necessary to accommodate the DNA superhelical path in the nucleosome. DNA segments bent into the minor groove are either kinked or alternately shifted. The unusual DNA conformational parameters induced by the binding of histone protein have implications for sequence-dependent protein recognition and nucleosome positioning and mobility. Comparison of the 147-base-pair structure with two 146-base-pair structures reveals alterations in DNA twist that are evidently common in bulk chromatin, and which are of probable importance for chromatin fibre formation and chromatin remodelling.

DNA in eukaryotic cells is packaged repetitively into nucleosomes by means of extensive association with histone proteins^{1–3}. The hierarchical chromatin structure formed is the genomic substrate relevant to the vital processes of DNA replication, recombination, transcription, repair and chromosome segregation, and to the pathological progression of cancer and viral disease. Although nucleosomal organization of DNA is essentially ubiquitous throughout genomes and generally repressive to gene expression, it also contributes to gene transcription in a gene-specific manner, suggesting that nucleosome positioning in gene promoter regions is important for genuine gene regulation *in vivo*^{4–6}. The question therefore arises of how chromatin structure, in which DNA is normally highly compacted, permits site-specific access to regulatory factors and more extensive exposure to the transcription apparatus.

The answer is likely to require a knowledge of DNA conformation in the nucleosome core. The core comprises 147 base pairs (bp) of DNA and the histone octamer; compared with the nucleosome, it lacks only 10–90 bp of linker DNA envisaged to be naked or bound to histone H1. The histone-fold domains of the octamer organize the central 129 of 147 bp in 1.59 left-handed superhelical turns with a diameter only fourfold that of the double helix. The relatively straight 9-bp terminal segments contribute little to the curvature of the complete 1.67-turn superhelix⁷. So far, the site-specific regulatory factors that have been discovered bind the linker or terminal regions of the intact nucleosome. The lack of binding to the central region of the superhelix might simply be a consequence of bending the double helix or, additionally, of unusual DNA conformations induced by histone binding. At least one protein, HIV-1 integrase, does prefer DNA bent around the nucleosome in contrast to naked DNA⁸.

The initiation of DNA-dependent nuclear processes in the context of chromatin implies that nucleosome position is biased by the DNA sequence to facilitate access by initiation factors. Numerous examples of positioned nucleosomes in gene promoter regions have been described both *in vivo* and *in vitro*^{9–11}. Preferential positioning could place factor-binding sequences in nucleosome linker or terminal region DNA. Furthermore, nucleosomes are intrinsically mobile and yield access to their DNA *in vitro*, allowing even RNA polymerase to transcribe nucleosomal DNA without causing dissociation of the histone octamer^{12–14}. *In vivo*, energy-dependent chromatin remodelling factors, targeted by gene regulatory proteins and acting directly on the nucleosome core, augment nucleosome

mobility¹⁵. Their mechanism of action most probably derives from the innate ability of nucleosomes to ‘slide’ along DNA without releasing it³.

An accurate, atomic-level description of DNA conformation in the nucleosome core and comparison with naked DNA is essential to an understanding of chromatin properties such as those resulting from nucleosome position and mobility. Previously, the probable sequence-dependent features of nucleosomal DNA were proposed by extrapolation from high-resolution DNA oligonucleotide structures¹⁶. Here, a direct analysis of nucleosome DNA is made, permitted by the high quality of the DNA structure in the 1.9-Å-resolution X-ray structure of the nucleosome core particle (NCP147)^{17,18}, comparable to high-resolution oligonucleotide structures. The persistently tight curvature of nucleosome core DNA causes it to take on conformations not obviously present in oligonucleotides but about which there has been speculation for many years¹⁹. Furthermore, comparison of the DNA twist values measured for the 147-bp and 146-bp structures with the DNA sequence periodicity found in bulk chromatin indicates that the DNA stretching observed in the nucleosome core is commonplace *in vivo*.

Excess DNA curvature

The ideal DNA superhelix that best fits the NCP147 DNA was constructed from uniformly distributed base pairs by using B-form DNA geometry. It has a radius of 41.9 Å and a pitch of 25.9 Å (Fig. 1a). Each base-pair step comprising two adjacent base pairs contributes 4.53° to the curvature of the ideal superhelix. However, the NCP147 superhelix is not bent uniformly owing to (1) the anisotropic flexibility of DNA, (2) local structural features intrinsic to the DNA sequence, and (3) irregularities dictated by the underlying histone octamer. The NCP147 DNA is generally B-form as judged by the criterion of the phosphate ‘Z-coordinate’ with a value of -0.18 ± 0.69 Å (mean $\pm$ s.d.; range -2.32 to 1.45 Å). On the basis of oligonucleotide structures, values falling in the range from -1 to 0 are indicative of the B-form, whereas values above 2 signify the A-form²⁰. Other standard conformational parameters also indicate that the DNA is predominantly B-form (see Supplementary Table 1).

Remarkably, the NCP147 DNA has double the base-pair-step curvature necessary to produce the DNA superhelix path. The ideal superhelix fit to the NCP147 DNA yields 1.67 superhelical turns for 133.6 bp (1.84 superhelical turns for 147 bp), bending the DNA

through 600.0°, whereas the actual curvature of NCP147 DNA is 1,333.3° (both scalar curvature values were calculated with the program Curves²¹). The use of only the central 129bp to remove the effect of straightening of the superhelical path at the termini yields values of 579.3° and 1248.6° ($9.75 \pm 5.13^\circ$ per base-pair step)

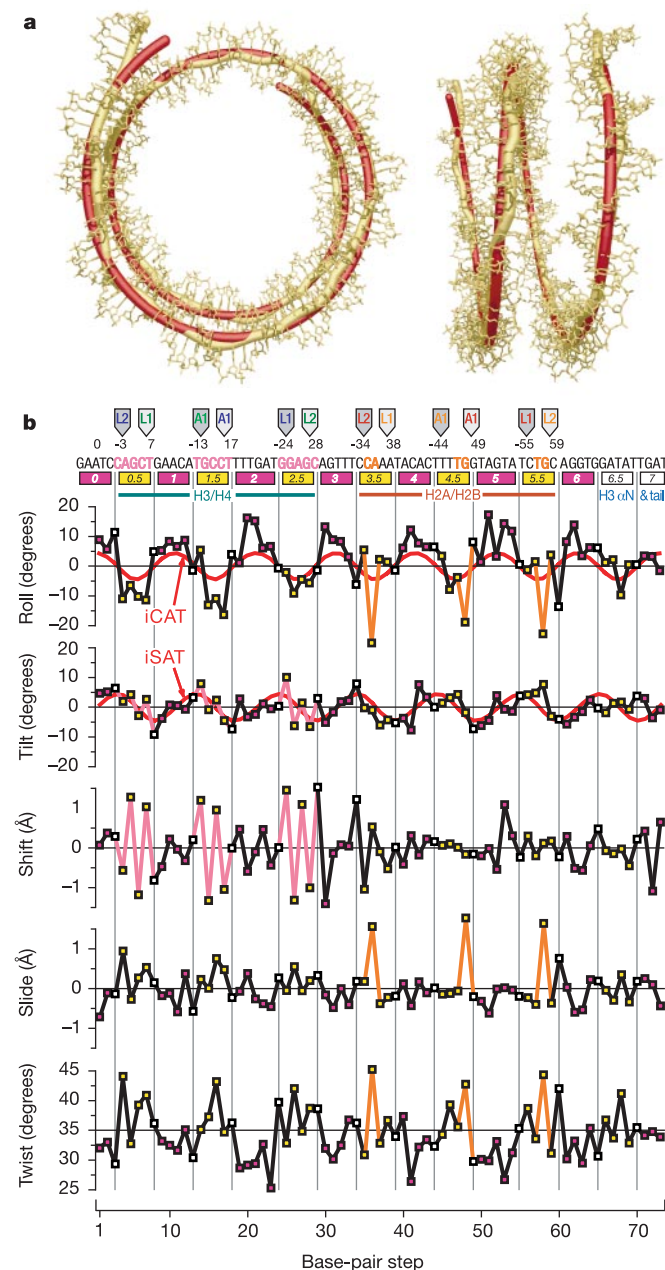


Figure 1 Superhelical path and base-pair-step parameters for NCP147 DNA. **a**, Structural alignment of the NCP147 (gold) and best-fit, ideal superhelix (red) paths. The NCP147 DNA structure is superimposed (gold). The left view is down the superhelix axis and the right view is rotated 90° around the pseudo-two-fold axis (vertical). **b**, Roll, tilt, shift, slide and twist base-pair-step parameters. The base-pair-step values plotted are the means for the two halves of the symmetrical sequence (shown above, with the dyad position labelled '0'). The iCAT and iSAT curves (red) show the roll and tilt contributions to the ideal superhelix curvature. The minor-groove blocks show base-pair-step shift alternation (pink in shift and tilt) or kinks (orange in roll, slide and twist). The primary bound-phosphate groups are indicated above the base sequence by pointers (numbered as the 5' phosphate of the adjacent base) showing the strand direction (dark, 3' → 5'; light, 5' → 3') and the interacting histone motif (L1, loop 1; L2, loop 2; A1, α1).

for the ideal and actual curvatures, respectively, and thus the terminal regions do not account for the discrepancy. The excess is not adequately explained by the zigzag path of the NCP147 superhelix, as the overall root-mean-square deviation from the ideal path is only 1.42 Å (Fig. 1a). We find that the greater part of the excess curvature is a consequence of alternation of DNA form not revealed by criteria based on oligonucleotide structures. Determining the component of curvature directed towards superhelix formation exposes the origin of the form changes.

The potential contributions to curvature from the roll and tilt (see Supplementary Fig. 1) of each base-pair step vary as sinusoidal functions of the accumulated twist angle (Θ) along the length of the double helix. To evaluate the actual contributions, the roll and tilt values for each base-pair step are multiplied by the corresponding $\cos \Theta$ (CAT) and $\sin \Theta$ (SAT), respectively. A convenient origin for the summation of Θ is taken as the central base pair located on the pseudo-two-fold symmetry axis of the particle (base pair 0 or dyad position). For the ideal superhelix, the CAT and SAT values multiplied by 4.53° yield the roll and tilt contributions to superhelix curvature at each base-pair step (iCAT and iSAT; Fig. 1b). The actual roll and tilt values for NCP147 DNA are moderately well correlated with CAT ($R = 0.72$, $P < 0.0001$) and SAT ($R = 0.54$, $P < 0.0001$), respectively. The sum of roll and tilt multiplied by CAT and SAT, respectively, over all base-pair steps yields a curvature free of the effect of path deviations between the ideal and actual superhelices. The resulting value is 897.0°, or 1.5-fold that required to accommodate the full NCP147 superhelix.

Unlike the computer-generated ideal superhelix, for which curvature derives equally from roll and tilt (iCAT and iSAT; Fig. 1b), curvature for oligonucleotide and NCP147 DNA stems predominantly from the roll parameter. For DNA oligonucleotide structures, roll is favoured over tilt by 2:1 (ref. 22), presumably because it requires less base-pair unstacking for any particular bend angle. In the nucleosome core, this ratio is virtually identical at 1.9:1, but when the components that contribute to the superhelix only are considered, the ratio is 3.0:1. Curvature into the superhelix is therefore accumulated every approximately five base pairs, where either the major or minor groove faces the histone octamer. Nevertheless, the NCP147 tilt component contributes 77% of that for the ideal superhelix (Fig. 1b; compare tilt and iSAT plots). In contrast, the roll component is 222% of the ideal contribution (Fig. 1b; compare roll and iCAT plots).

Analyses of crystal structures of DNA oligonucleotides and protein–DNA complexes normally equate DNA bending with DNA curvature²³ and rely on base-pair-step parameters to explain bending behaviour; for example, by using roll and tilt^{22,24}, roll and twist^{16,25,26} or roll, tilt, twist, slide and rise^{24,27}. This type of approach is inadequate for nucleosome core DNA, given the substantial excess curvature that contributes to superhelix formation. Further analysis indicates that the conformation of nucleosome core DNA is not typical of oligonucleotide DNA and has not previously been described in protein–DNA complexes. Principal-component analysis of DNA structural parameters of both nucleosome core and oligonucleotide DNA show that the two most significant base-pair-step parameter correlations are roll–slide–twist and tilt–shift (see Supplementary Table 2). However, the coefficients for the base-pair-step components are significantly larger for NCP147 DNA (mean magnitudes are 0.88 versus 0.57 for roll–slide–twist and 0.86 versus 0.69 for tilt–shift), indicating that the NCP147 DNA is limited to a smaller regime of conformational space than is occupied by oligonucleotide DNA. When the backbone angles are also included in the analysis, the ϵ , ζ and β angles are found tightly linked to the roll–slide–twist component. Stronger parameter coupling occurs for blocks of base pairs with the minor groove as opposed to the major groove facing the histone octamer (mean magnitudes of 0.69 versus 0.27 for roll–slide–twist with backbone angles included), indicating differences in modes of

DNA bending or in conformational variability for the two orientations.

Major-groove versus minor-groove curvature

The DNA curvature of the central 129 bp of NCP147 favours the major groove over the minor groove by 1.3:1 per base-pair step in contrast to the approximate 2:1 preference observed for oligonucleotide DNA (evaluated from data in ref. 20). Minor-groove bending is facilitated in NCP147 by the insertion of an arginine side chain into the minor groove at each of the 14 sites at which it faces the histone octamer. The minor groove narrows to a width of $3.0 \pm 0.55 \text{ \AA}$ at these sites¹⁸. An insightful analysis is made by dividing the DNA along its length into blocks of continuous positive and negative CAT values, defining respectively regions where the major-groove and minor-groove curvatures contribute to superhelix formation (Fig. 1b). A single, junction base-pair step joins the major-groove and minor-groove blocks and cannot contribute significantly to superhelix curvature by base-pair-step roll because these steps have the minimum magnitudes for CAT (-0.10 to 0.26). The major-groove blocks display 'smooth' bending with all roll angles positive ($8.47 \pm 5.18^\circ$) and with systematic underwinding of the DNA indicated by the reduced twist angles and negative slide displacements (Figs 1b and 2a). The minor-groove blocks exhibit 'smooth' bending only over the H3–H4 tetramer, with essentially all base-pair steps having negative roll angles ($-7.86 \pm 6.02^\circ$). In contrast, minor-groove blocks are kinked over the H2A–H2B dimers ($-5.13 \pm 11.01^\circ$; Fig. 1b). This difference in minor-groove-block bending modes associated with H3–H4 and H2A–H2B is most probably a consequence of specific DNA sequences and not underlying differences in the histone binding or degree of curvature (mean values over all base-pair steps: H3–H4, 9.8° ; H2A–H2B, 9.9°). Smooth bending in minor-groove blocks is associated with large alternation of shift values roughly between -1 and $+1 \text{ \AA}$ for four base-pair steps, which relieves steric interference between the base edges (Figs 1b and 2b). This mode of bending, which is the prime contributor to the observed shift–tilt coupling, depends most consistently on the GC base-pair step, and on GG = CC or AG = CT base-pair steps, with one and two in each block, respectively. GC steps have been noted previously for their high shift propensity²⁰. The GG = CC and AG = CT steps are, of all steps, the only two that can form cross-chain hydrogen bonds in the minor groove as observed for oligonucleotides²⁸ and for NCP147. The occurrence of two GG = CC/AG = CT base-pair steps in the minor-groove blocks bound to the H3–H4 tetramer, but not the

H2A–H2B dimer, is a probable determinant of smooth versus kinked bending.

Kinking in minor-groove blocks always occurs at a single CA = TG base-pair step that has a roll angle in the range -18° to -27° , roughly the total curvature into the superhelix for the entire block (Figs 1b and 2c). These kinked steps have a concomitant large slide value of over $1.5 \text{ \AA}$ and are also overtwisted to values larger than 40° . They have a conformation that is extreme in roll–slide–twist coupling. Ten of 12 minor-groove blocks have one CA = TG step and two have none. The six CA = TG steps that are kinked have CAT values between -0.6 and -1 , whereas the four non-kinked CA = TG steps have values more positive than -0.5 . As steps with more negative CAT values have more potential to contribute to superhelix curvature, they are conversely under more stress to kink. The occurrence of CA = TG steps in kinks is consistent with their properties as gleaned from oligonucleotide structures²⁰: (1) TA and CA = TG steps are the most flexible as judged from the standard deviation of roll angles, (2) the mean slide for CA = TG is $1.18 \text{ \AA}$ —the only base-pair step with a value greater than $1 \text{ \AA}$ (the TA mean value is $-0.80 \text{ \AA}$), and (3) with a mean twist angle of 37.4° , CA = TG is one of the most overwound base-pair steps. TA and CA = TG are by far the most flexible of the 10 base-pair steps with regard to roll, slide and twist, but the propensity of TA for negative slide makes it much less likely to kink in the manner observed in NCP147 DNA. Clash of the purines across the minor groove in the kinks observed is avoided because of large slide values. Furthermore, TA steps have the largest positive mean roll angle (11.8° versus the next largest of 6.3° for GG = CC), which is consistent with their appearance only in major-groove blocks of NCP147. The importance of base-pair-step flexibility in nucleosome positioning and stability has been suggested previously^{29,30}. A survey of protein–DNA complexes shows that only seven base-pair steps are rolled into the minor groove to a degree comparable to that of NCP147 DNA, but none of these is a CA = TG step³¹. Comparison of the mean roll angle for each type of base-pair step within NCP147 DNA and from oligonucleotide structures²⁰ shows that they are only weakly correlated with each other (see Supplementary Information).

DNA form and superhelix formation

Strikingly for NCP147 DNA, the conformational parameter tip oscillates regularly along the length of the DNA (Fig. 3a). The transitions between negative and positive tip angles represent an alternation of the DNA form that manifests itself in the major and minor blocks as excess roll. By definition, the excess roll of a series of

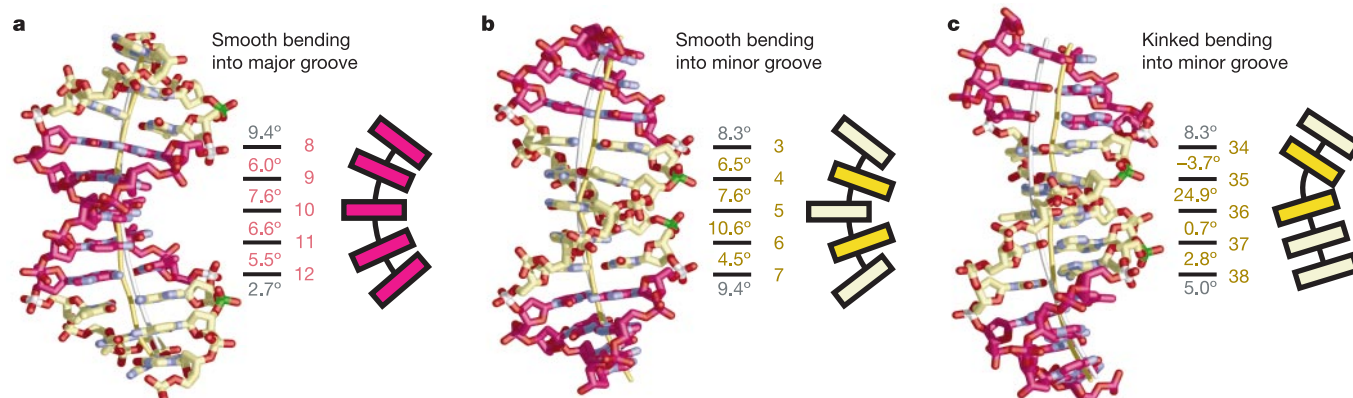


Figure 2 DNA bending in the NCP147 DNA. Structures (left) and schematic representations (right) stress uniformity of curvature in the major-groove blocks (red) (a), and alternating shift values (b) and localization of curvature in kinks in minor-groove blocks (c) (yellow for one representative double-helical turn). Also indicated are the

primary bound-phosphate groups (green), the block-junction phosphate groups (white) and the DNA axes for the NCP147 (gold) and ideal (white) superhelices. The contributions of base-pair-step curvature to superhelix bending are listed with base-pair numbers (centre).

base-pair steps is the sum of the change in tip angles over the corresponding base pairs (Fig. 3b). The source of excess roll in the NCP147 superhelix is localized primarily to block junctions and the base pairs immediately adjacent on either side, where the tip is significantly non-zero. The reason for the increased tip-angle magnitudes at these sites is that they accommodate the smaller radius of curvature on the inside surface of the superhelix compared with the outside, about 32 Å compared with 52 Å. In principle, this radius-of-curvature difference should require shorter phosphate-phosphate distances (PP) where the DNA backbone faces the histones than where it faces away. In fact, it is not the absolute distance that is crucial but the component of the PP distance (PP_a) lying parallel to the path of the superhelix. As two adjacent base-pair steps were found compensatory for this distance, PP_a is obtained from three base pairs ($i - 1$ to $i + 1$), and the direction of the superhelix path is defined by four points on the double-helix axis ($i - 1$ to $i + 2$). Importantly, it is PP_a that is shortened on the inside and lengthened on the outside of the DNA superhelix by the oscillating tip angles. The difference calculated at each base pair between the values of PP_a for the two phosphodiester chains (ΔPP_a) is strongly correlated with tip angle ($R = 0.91$, $P < 0.0001$; Fig. 3a). The values of ΔPP_a compared with SAT, a direct measure of inside and outside, are somewhat less well correlated ($R = 0.83$, $P < 0.0001$) owing to DNA-sequence-specific effects such as kinking in the minor-groove blocks.

Phosphate-phosphate distances, as opposed to their superhelical path component, are not correlated with SAT, curvature or roll angles, and are only weakly correlated with tip. This is the expected behaviour because the phosphodiester backbone is essentially incompressible, as described previously for oligonucleotide structures³², having in NCP147 a mean phosphate-phosphate distance of 6.68 ± 0.23 Å between nearest neighbours along the same chain. However, a notable variation in the B-form backbone is the anticorrelation of the ϵ and ζ backbone angles giving rise to the B_I and B_{II} forms³³. In NCP147, the population of base-pair steps intermediate between B_I and B_{II} (that is, $B_{I/II}$) is larger than for oligonucleotides, but the distribution is not significantly correlated with other form or backbone parameters. The differences in mean phosphate-phosphate distances for B_I , $B_{I/II}$ and B_{II} forms separately, even taking into consideration the location in major-groove and minor-groove blocks or junctions, are less than 1 s.d. of the mean distance overall. Nevertheless, in minor-groove blocks, a B_{II} to B_I transition occurs at kinks, and B_{II} and B_I alternation occurs at sites of alternating shift. In the latter case, the B_{II} base pair is displaced into the minor groove and has up to -40° of propeller twist. This unusual conformation for GG and AG base-pair steps is stabilized by cross-strand bifurcated hydrogen bonds in the minor groove between guanine N2 and pyrimidine O2 atoms. Moreover, the variability observed between repeating binding sites seems to stem from a dependence on DNA sequence, not only for the choice of

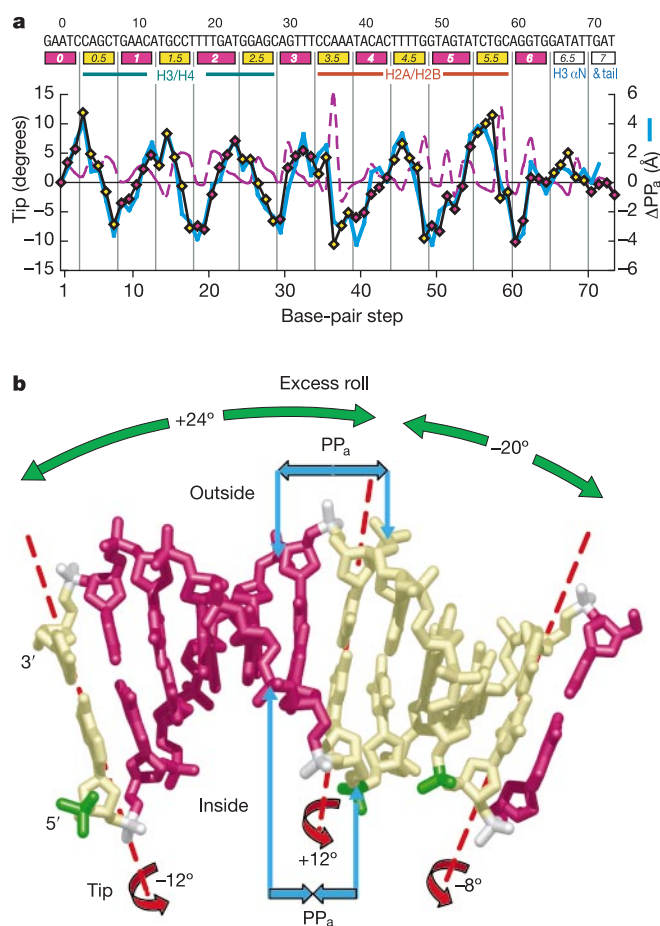


Figure 3 Oscillation of the base-pair-tip parameter for NCP147 DNA. **a**, Correlation of base-pair tip and ΔPP_a , plotted as in Fig. 1b. Excess base-pair-step roll (purple) into the superhelix is a consequence of oscillating base-pair-tip angles (black). ΔPP_a (blue) is the difference between PP_a values for the two phosphodiester chains at a single base pair. **b**, Coupling between PP_a , tip and excess roll. PP_a is the component of the phosphate-

phosphate distance that lies in the plane containing the local superhelical path. Rotation of a base pair along its tip axis (red) in the direction indicated decreases PP_a on the inside and increases it on the outside of the superhelix. Excess roll for a base-pair step is the difference between the tip angles for the two contributing base pairs. The DNA blocks shown are from SHL0.0 and SHL0.5.

smooth or kinked bending in minor-groove blocks but also for the precise details of the conformation along the entire DNA.

Histone constraints on DNA conformation

Conformational differences between nucleosome core and oligo-nucleotide DNA are probably important for the recognition, or lack of it, of nucleosome DNA by nuclear factors, and for the folding of nucleosome arrays into the chromatin fibre. Furthermore, the dependence of nucleosome position and stability on base sequence, as demonstrated convincingly by experiments *in vitro* (for example ref. 11), derives from the energetics of sequence-dependent histone–DNA interactions. Parameterization of an energy function, analogous to that for protein–DNA complexes³⁴, capable of accurate prediction of sequence-dependent behaviour will require many high-resolution structures of the nucleosome core particle containing different DNA sequences. Currently, comparisons between histones H2A, H2B, H3 and H4 for the three primary types of DNA-binding motif in the nucleosome core give a preliminary indication of the significance of histone restraints on DNA conformational flexibility.

The principal DNA-binding sites in the nucleosome core particle are the eight L1L2 loop and four $\alpha 1\alpha 1$ helix structures within the histone-fold domains⁷. Although the histone–DNA interactions at each of these sites are complex (a complete tabulation is given in ref. 17), one phosphate group from each DNA strand shows at least one conserved interaction with protein over the four types of core histone (Figs 1b and 4). These primary bound-phosphate groups are 5' to the last base pair for both DNA strands in each minor-groove block, with one exception at SHL4.5 (SHL: superhelix location⁷), resulting from a dissimilar packing arrangement of the $\alpha 1$ and $\alpha 2$ helices for H2B in comparison with the other histones⁷. Separate structural alignments, calculated with the histone main-chain segments only, for the L1, L2 and A1 binding motifs (the components of L1L2 loop and $\alpha 1\alpha 1$ helix sites) and their associated phosphate groups show the precision with which the selected phosphate groups are localized by binding to histone (see Supplementary Table 3).

For the L1 and L2 binding motifs, the root-mean-square deviations for the primary bound-phosphate group (1.27 and 1.39 Å) and its 3'-adjacent neighbour (1.42 and 1.48 Å) are strikingly low

relative to the mean phosphate–phosphate distance of 6.68 Å (see Supplementary Table 3). The values for the phosphate groups are comparable to those for the segments of protein main-chain (0.75 and 0.46 Å) used for the alignments if the phosphates associated with the structurally deviant H2B are not included. Two phosphate groups separated by 21 bases along each DNA strand are therefore well localized on both the H3–H4 and H2A–H2B histone-fold pairs (for example H4:L1 to H4:L2) through histone interaction. Correspondingly, the DNA segments bridging between the halves of the H3–H4 tetramer and H3–H4 tetramer to H2A–H2B dimer are 10 bases in length. Evidently, the histone proteins impose substantial conformational restraints on the DNA, which responds in a sequence-dependent manner, such as the smooth and kinked bending modes observed for the minor-groove blocks. The phosphate deviations for the A1 motif are larger than for L1 and L2 motifs, which might reflect their location between phosphate groups at 10.5 bases. This non-optimal arrangement might decrease sequence-dependent effects on stability by accommodating either 10 or 11 bases between A1 motifs and the adjacent L1 and L2 motifs. A possible example of this multiplicity occurs for H2B A1 at the SHL4.5/5.0 block junction (Fig. 1b). Indeed, the location of A1 relative to L1 and L2 motifs could promote DNA stretching.

DNA stretching

The double-helical twist values for NCP147 DNA range from 23.7° to 47.9° and have a mean value of $34.52 \pm 4.75^\circ$, or equivalently 10.43 bp per turn, which corresponds to the laboratory reference frame³⁵. Values from experiments that probe DNA by enzymatic digestion or hydroxyl radical cleavage correspond to the local reference frame and are calculated from the laboratory frame by removing the geometrical contribution from the superhelix pitch. Assuming a uniform pitch angle α , then $\text{Twist}(\text{local}) = \text{Twist}(\text{laboratory}) - (2\pi \sin \alpha)/N$, where N is the number of base-pair steps in one superhelical turn. Using the values from the ideal superhelix of -5.62° and 78.90 steps for α and N , respectively, NCP147 overall has a local frame twist of 10.30 bp per turn. Remarkably, two 146-bp nucleosome core particle structures containing differing DNA sequences have values of 10.23 and 10.15 bp per turn¹⁷. These reduced values are a consequence of DNA stretching by 1–2 bp to satisfy the crystal packing constraints of

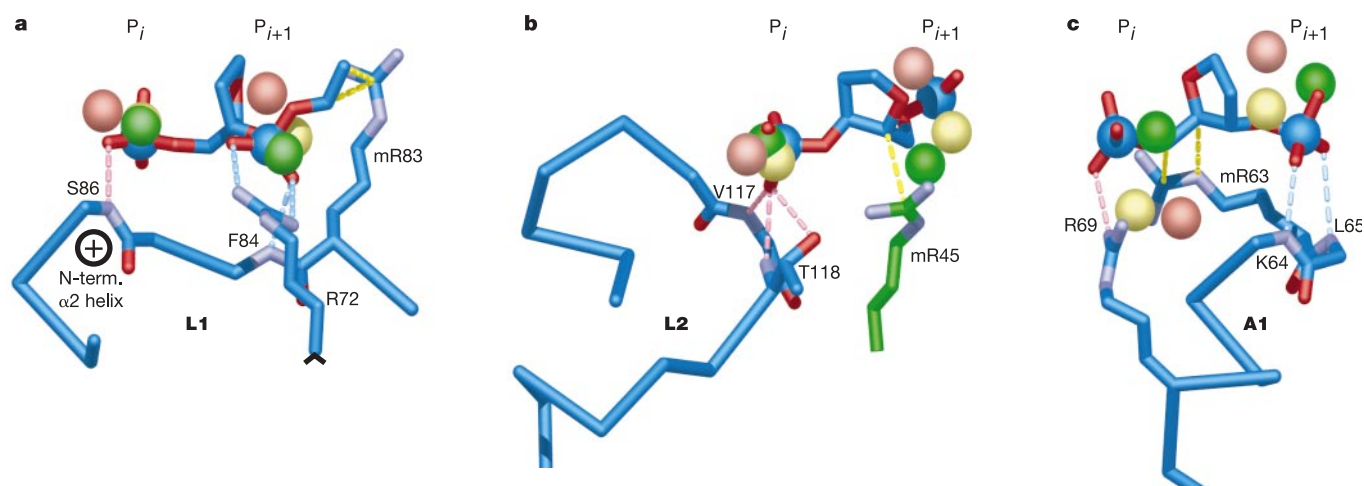


Figure 4 Structural alignments of the histone-fold DNA-binding motifs and bound-phosphate groups. The motifs are L1 (**a**), L2 (**b**) and A1 (**c**). For each, the main-chain C_{α} atoms of H4, H2A and H2B were aligned with those for H3. The resulting primary bound-phosphate group (P_i) and 3'-adjacent phosphate group (P_{i+1}) positions are shown (H3,

blue; H4, green; H2A, yellow; H2B, red) with respect to the H3 main-chain and interacting groups. Completely conserved (pink) and partly conserved (cyan) hydrogen bonds and interactions (yellow) with the minor-groove arginine side chain (mR) are shown.

the DNA termini⁷. The actual 12-bp region stretched is in a different location in the two 146-bp structures, and both regions are distant from the DNA termini. Evidently, stretching at the DNA termini introduces a twist defect in the double helix that then diffuses to predominantly one region in each particle, depending on the DNA sequence. The distorted region is centred on either an H3–H4 or an H2A–H2B α 1 α 1 DNA-binding site and is bound at each end by the adjacent L1L2 sites. These two 146-bp structures therefore contain DNA regions that represent trapped intermediates relevant to a twist-defect-diffusion mechanism of nucleosome mobility^{36,37} and twist propagation by chromatin-remodelling enzymes.

DNA isolated from nucleosome core particles prepared from bulk-source endogenous chromatin have a sequence periodicity of 10.17 bp per turn³⁵, which is significantly closer to the local frame values for the 146-bp versus 147-bp nucleosome core particles. This similarity suggests that nucleosomes in the cell nucleus have evolved to contain DNA stretched on average by 1–2 bp. The difference between 10.17 bp per turn for nucleosomes and 10.5 bp per turn, accepted for the average twist of naked DNA³⁵, accounts for the discrepancy in the observed and effective values (–1.67 versus –1.2 (ref. 38)) of nucleosome supercoiling. The overtwisting of the DNA double helix on the nucleosome, due in part to DNA stretching, resolves the ‘linking number paradox’³⁹ observed for nucleosomes assembled on closed circular DNA. Unusual twisting or supercoiling of the linker DNA between nucleosome cores in these arrays is not required. Formation of a compact nucleosome higher-order structure could be facilitated by DNA stretching because it would provide a buffer against incompatible DNA linker lengths. Each linker DNA segment running between two nucleosome cores could on average adjust in length by up to four base pairs by stretching or compressing the DNA bound to the neighbouring H2A–H2B and H3–H4 histone-fold domains. Not only would this provide for linker length adjustment, it would also alleviate twist angle restrictions of at least 140°. □

Methods

Crystal preparation, data collection, model refinement and figure preparation were as described previously¹⁸. The ideal DNA superhelix was calculated by using B-form base-pair geometry⁴⁰ and fitted by least squares to the NCP147 DNA (T.J.R., unpublished observations). The twist offset for the central base pair of NCP147 is $\pm 4.99^\circ$ (negative for strand with A central). It is presumably non-zero because of the asymmetry of the AT base pair lying on the molecular dyad of the particle. DNA curvature, base-pair and base-pair-step parameters, and backbone geometry were calculated with the program Curves²¹ (see Supplementary Methods). Excel (Microsoft) and SigmaPlot (SPSS) were used for the analysis of DNA geometry. Correlation coefficients were calculated by using variance weighting. Principal components were calculated with IMSL routines (T.J.R., unpublished observations).

Received 16 January; accepted 12 March 2003; doi:10.1038/nature01595.

- van Holde, K. E. In *Chromatin* (ed. Rich, A.) (Springer, New York, 1988).
- Kornberg, R. D. & Lorch, Y. Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* **98**, 285–294 (1999).
- Elgin, S. C. R. & Workman, J. L. (eds) *Chromatin Structure and Gene Expression* (Oxford Univ. Press, Oxford, 2000).
- Simpson, R. T. Nucleosome positioning: Occurrence, mechanisms, and functional consequences. *Prog. Nucleic Acid Res. Mol. Biol.* **40**, 143–184 (1991).
- Wolffe, A. P. & Kurumizaka, H. The nucleosome: A powerful regulator of transcription. *Prog. Nucleic Acid Res. Mol. Biol.* **61**, 379–422 (1998).
- Wyrick, J. J. et al. Chromosomal landscape of nucleosome-dependent gene expression and silencing in yeast. *Nature* **402**, 418–421 (1999).
- Luger, K., Maeder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
- Pryciak, P. M. & Varmus, H. E. Nucleosomes, DNA-binding proteins, and DNA sequence modulate retroviral integration target site selection. *Cell* **69**, 769–780 (1992).
- Kornberg, R. D. & Lorch, Y. Chromatin structure and transcription. *Annu. Rev. Cell Biol.* **8**, 563–587 (1992).
- Shimizu, M., Roth, S. Y., Szent-Gyorgyi, C. & Simpson, R. T. Nucleosomes are positioned with base pair

- precision adjacent to the alpha 2 operator in *Saccharomyces cerevisiae*. *EMBO J.* **10**, 3033–3041 (1991).
- Flaus, A. & Richmond, T. J. Positioning and stability of nucleosomes on MMTV 3' LTR sequences. *J. Mol. Biol.* **275**, 427–441 (1998).
- Meersseman, G., Pennings, S. & Bradbury, E. M. Mobile nucleosomes—a general behaviour. *EMBO J.* **11**, 2951–2959 (1992).
- Polach, K. J. & Widom, J. Mechanism of protein access to specific DNA sequences in chromatin: A dynamic equilibrium model for gene regulation. *J. Mol. Biol.* **254**, 130–149 (1995).
- Studitsky, V. M., Kassavetis, G. A., Geiduschek, E. P. & Felsenfeld, G. Mechanism of transcription through the nucleosome by eukaryotic RNA polymerase. *Science* **278**, 1960–1963 (1997).
- Tsukiyama, T. The *in vivo* functions of ATP-dependent chromatin-remodelling factors. *Nature Rev. Mol. Cell Biol.* **3**, 422–429 (2002).
- Calladine, C. R. & Drew, H. R. Principles of sequence-dependent flexure of DNA. *J. Mol. Biol.* **192**, 907–918 (1986).
- Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).
- Davey, C. A. & Richmond, T. J. DNA-dependent divalent cation binding in the nucleosome core particle. *Proc. Natl Acad. Sci. USA* **99**, 11169–11175 (2002).
- Crick, F. H. C. & Klug, A. Kinky helix. *Nature* **255**, 530–533 (1975).
- El Hassan, M. A. & Calladine, C. R. Conformational characteristics of DNA: Empirical classifications and a hypothesis for the conformational behaviour of dinucleotide steps. *Phil. Trans. R. Soc. Lond. A* **355**, 43–100 (1997).
- Lavery, R. & Sklenar, H. The definition of generalised helicoidal parameters and of axis curvature for irregular nucleic acids. *J. Biomol. Struct. Dynam.* **6**, 63–91 (1988).
- Young, M. A., Ravishanker, G., Beveridge, D. L. & Berman, H. M. Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA–protein complexes. *Biophys. J.* **68**, 2454–2468 (1995).
- Olson, W. K. Simulating DNA at low resolution. *Curr. Opin. Struct. Biol.* **6**, 242–256 (1996).
- Dickerson, R. E. DNA bending: The prevalence of kinkiness and the virtues of normality. *Nucleic Acids Res.* **26**, 1906–1926 (1998).
- Goodsell, D. S. & Dickerson, R. E. Bending and curvature calculations in B-DNA. *Nucleic Acids Res.* **22**, 5497–5503 (1994).
- El Hassan, M. A. & Calladine, C. R. Two distinct modes of protein-induced bending in DNA. *J. Mol. Biol.* **282**, 331–343 (1998).
- Packer, M. J. & Hunter, C. A. Sequence–structure relationships in DNA oligomers: A computational approach. *J. Am. Chem. Soc.* **123**, 7399–7406 (2001).
- Yanagi, K., Prive, G. G. & Dickerson, R. E. Analysis of local helix geometry in three B-DNA decamers and eight dodecamers. *J. Mol. Biol.* **217**, 201–214 (1991).
- Shrader, T. E. & Crothers, D. M. Effects of DNA sequence and histone–histone interactions on nucleosome placement. *J. Mol. Biol.* **216**, 69–84 (1990).
- Anselmi, C., Bocchini, G., De Santis, P., Savino, M. & Scipioni, A. Dual role of DNA intrinsic curvature and flexibility in determining nucleosome stability. *J. Mol. Biol.* **286**, 1293–1301 (1999).
- Dickerson, R. E. & Chiu, T. K. Helix bending as a factor in protein/DNA recognition. *Biopolymers* **44**, 361–403 (1997).
- Packer, M. J. & Hunter, C. A. Sequence-dependent DNA structure: The role of the sugar-phosphate backbone. *J. Mol. Biol.* **280**, 407–420 (1998).
- Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.* **257**, 14686–14707 (1982).
- Olson, W. K., Gorin, A. A., Lu, X. J., Hock, L. M. & Zhurkin, V. B. DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl Acad. Sci. USA* **95**, 11163–11168 (1998).
- Travers, A. A. & Klug, A. The bending of DNA in nucleosomes and its wider implications. *Phil. Trans. R. Soc. Lond. B* **317**, 537–561 (1987).
- Richmond, T. J. & Widom, J. In *Chromatin Structure and Gene Expression* (eds Elgin, S. C. R. & Workman, J. L.) 1–23 (Oxford Univ. Press, Oxford, 2000).
- Widom, J. Role of DNA sequence in nucleosome stability and dynamics. *Q. Rev. Biophys.* **34**, 269–324 (2001).
- Zivanovic, Y., Goulet, I., Revet, B., Le Bret, M. & Prunell, A. Chromatin reconstitution on small DNA rings II. DNA supercoiling on the nucleosome. *J. Mol. Biol.* **200**, 267–290 (1988).
- Klug, A. & Lutter, L. C. The helical periodicity of DNA on the nucleosome. *Nucleic Acids Res.* **9**, 4267–4283 (1981).
- Arnot, S., Dover, S. D. & Wonacott, A. J. Least-squares refinement of the crystal and molecular structures of DNA and RNA from X-ray data and standard bond lengths and angles. *Acta Crystallogr. B* **25**, 2192–2206 (1969).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank I. Berger and D. Sargent for comments on the manuscript. This study was supported by the Swiss National Science Fund.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.J.R. (richmond@mol.biol.ethz.ch). The coordinates of the X-ray structures are available from the Protein Data Bank as files 1kx3, 1kx4 and 1kx5.

Photographic observations of Neuschwanstein, a second meteorite from the orbit of the Příbram chondrite

Pavel Spurný*, Jürgen Oberst† & Dieter Heinlein‡

* Ondřejov Observatory, Astronomical Institute of the Academy of Sciences of the Czech Republic, 251 65 Ondřejov, Czech Republic

† DLR Institute of Space Sensor Technology and Planetary Exploration, Rutherfordstraße 2, 12489 Berlin, Germany

‡ DLR Fireball Network, Lilienstraße 3, 86156 Augsburg, Germany

Photographic observations of meteoroids passing through the atmosphere provide information about the population of interplanetary bodies in the Earth's vicinity in the size range from 0.1 m to several metres. It is extremely rare that any of these meteoroids survives atmospheric entry to be recovered as a meteorite on the ground. Příbram was the first meteorite (an ordinary chondrite) with a photographically determined orbit; it fell on 7 April 1959 (ref. 1). Here we report the fourth meteorite fall to be captured by camera networks. We determined the atmospheric trajectory and pre-atmospheric orbit of the object from the photographic records. One 1.75-kg meteorite—named Neuschwanstein and classified as an enstatite chondrite²—was recovered within the predicted impact area. The bolide's heliocentric orbit is exceptional as it is almost identical to the orbit of Příbram, suggesting that we have discovered a 'stream' of

meteoritic objects in an Earth-crossing orbit. The chemical classifications and cosmic-ray exposure ages of the two meteorites are quite different, however, which implies a heterogeneous stream.

On 6 April 2002, a very bright fireball illuminated large regions of western Austria and southern Bavaria (the time of brightest flare was 20:20:17.7 UT). The fireball was observed by outdoor witnesses in almost all of central Europe. Residents in towns and villages near the end of bolide trajectory reported shaking ground, rattling windows, and sounds audible within at least 100 km. In addition to the casual observations, the bolide was recorded by cameras, radiometers, infrasound detectors, and local seismic arrays, making this fireball an unusually well documented event. Here we analyse data obtained by a systematic long-term observational photographic programme—the European Fireball Network.

Photographic records were crucial for the determination of the fireball atmospheric trajectory (Fig. 1), the prediction of the meteorite impact area, and the derivation of the heliocentric orbit. Though the fireball was close to the horizon at all stations, the atmospheric trajectory could be determined by means of standard procedures^{3,4} with good precision (Table 1). The fireball's luminous trajectory (which was almost 91 km long) started at an altitude of 85.0 km, about 10 km east-northeast of Innsbruck, Austria, and terminated at an altitude of only 16.0 km, about 20 km west of Garmisch-Partenkirchen, Germany. The angle of the atmospheric trajectory to the Earth's surface was 49.5°. The maximum brightness (−17.2 absolute magnitude) was reached in a 0.1-s flare at a height of 21 km, probably marking a break-up of the meteoroid into several fragments. The fireball entered the atmosphere with a velocity of 20.95 km s^{−1} and decelerated to 2.4 km s^{−1}, when the ablation process stopped. The dynamic behaviour of this fireball in the atmosphere is of type I (ref. 5), which is usually associated with stony material.

The most probable value of the initial dynamic mass of the meteoroid entering the atmosphere is 300 kg, with an uncertainty of 100 kg. Only a small portion of about 20 kg, in several fragments, is expected to have reached the ground. Unfortunately, the individual meteoroid fragments cannot be directly observed in the photographic records. The unknown velocity increments following the break-up (which can be of the order of hundreds of m s^{−1}; ref. 6) therefore affect the uncertainty of the positions of the computed impact points. In addition, the dark flight of the meteorite fragments was significantly affected by wind. The predicted impact area of all fragments is approximately rectangular, being several kilometres long and about 800 m wide. Large fragments, less affected by atmospheric drag, are expected to be found downtrack of this target area. Though the predicted impact area in rugged mountain terrain appeared unsuited for search activities, a 1.75-kg stony meteorite (Fig. 2), classified as an EL6 enstatite chondrite², was found on 14 July 2002, about 400 m south of the computed trajectory (at 10°48.5' E, 47°31.5' N, altitude 1,650 m), about 6 km from the well-known castle Neuschwanstein. The meteorite was therefore named Neuschwanstein.

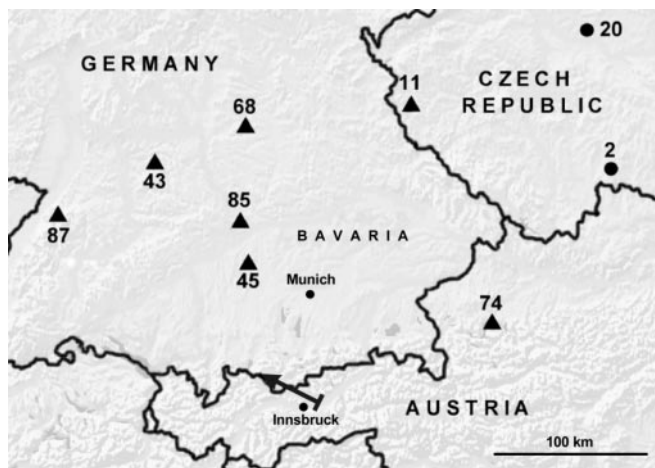


Figure 1 Map of part of central Europe, showing the projection of the Neuschwanstein atmospheric trajectory on the Earth's surface and the position of the optical instruments of the European Fireball Network. Numbered triangles show locations of the photographic cameras, as follows: 11, Přimda; 43, Öhringen; 45, Streitheim; 68, Losaurach; 85, Tüftstätt; 87, Gernsbach; 74, Gahberg. Large, numbered dots show locations of radiometers, as follows: 2, Kunžak; and 20, Ondřejov. The radiometers provided light-curve information and the exact time of the event. The German and Austrian stations of the European Fireball Network are equipped with mirror all-sky cameras, whereas the Czech stations have more accurate fish-eye cameras. Each of these stations takes one exposure of the whole sky per night. The exposure is interrupted by a rotating shutter, resulting in time marks on the photographic trails for velocity measurements. The network consists of about 30 stations and covers an area of approximately 10⁶ km². Among the photographic fireball networks that have operated in Europe (European Fireball Network^{18,19}, EN), in the USA (Prairie Network²⁰, PN) and in Canada (Meteorite Observation and Recovery Project²¹, MORP), only the EN is still currently active.

Table 1 Atmospheric trajectory data of the Neuschwanstein meteorite fall

	Beginning	Terminal
Velocity	20.95 ± 0.04 km s ^{−1}	2.4 ± 0.9 km s ^{−1}
Height	84.95 ± 0.04 km	16.04 ± 0.03 km
Longitude (E)	11.5524° ± 0.0009°	10.8507° ± 0.0008°
Latitude (N)	47.3039° ± 0.0006°	47.5257° ± 0.0005°
Dynamic mass	300 ± 100 kg	7 ± 3 kg*
Slope	49.75° ± 0.03°	49.23° ± 0.03°
Maximum abs. magnitude		−17.2
Total length/duration		90.6 km/5.3 s

* Terminal mass of the main piece; total mass of all fallen fragments is estimated at about 20 kg. All geographic coordinates are given using the WGS84 coordinate system.

We computed the heliocentric orbit of Neuschwanstein from the exact time of the fireball, its initial velocity, and the position of its radiant (Table 2). Before Neuschwanstein collided with Earth, it body moved around the Sun on an elliptic low-inclination orbit with the aphelion in the main belt of asteroids, which is quite common for fireballs that produce meteorites. However, the orbit of this bolide is exceptional, as it is practically identical to the orbit of the Příbram H5 ordinary chondrite, which fell on 7 April 1959 (Table 2, Fig. 3). The similarity of two orbits is typically described by the parameter D' . By definition⁷, two orbits are related when D' is smaller than 0.105. With $D' = 0.009$, the orbits of Příbram and Neuschwanstein are indeed remarkably similar.

Taking 200 sets of orbital elements of established 'meteorite candidates' as a basis⁸, we estimate that the chance of finding two meteorites among these 200 with orbits matching as well as those of Příbram and Neuschwanstein is one in 100,000. Therefore, this paired meteorite fall is probably not a coincidence, and implies that the two recovered samples are members of a 'stream' of similar objects. The existence of such meteorite streams was previously postulated on the basis of observed grouping of fireballs^{9,10}, Earth-crossing asteroids^{11,12} and historical meteorite falls^{13,14}. From the known fireball detection efficiency of the European Fireball Network (the cameras have monitored an area of 10^6 km^2 , 3 hours per day over the past 40 years, on average), and the separation of the two meteorite orbits in space, we estimate that the stream should contain about 10^9 meteoroids of comparable size. Assuming that this stream originated from the break-up of a single parent body, we calculate that this parent would have a minimum radius of 300 m. However, in the light of the different classifications of the two meteorites (Příbram is an H5 ordinary chondrite, Neuschwanstein an EL6 enstatite chondrite), a common primordial parent body for the two objects is most unlikely.

Also, the cosmic-ray exposure ages of Příbram and Neuschwanstein differ significantly—12 Myr (ref. 15) and 48 Myr (ref. 2), respectively. Both these ages, which are believed to represent the time span from the release of the meteoroid from the interior of a parent body to its encounter with Earth, are significantly longer than the typical survival times of meteoric streams.

If there is a stream of meteorite-producing bodies in the Příbram orbit, it may be worth searching for other members of this stream,



Figure 2 The Neuschwanstein meteorite. It is an EL6 enstatite chondrite with an original weight of 1.75 kg, but a 45-g sample has been cut off for analysis. Enstatite chondrites differ from ordinary chondrites in being more reduced, and having lower Mg/Si and (Ca,Al,Ti)/Si ratios. Meteorites with reliably determined orbits are very rare; all previous cases were ordinary chondrites.

Table 2 Radiants and orbital elements (epoch 2000.0)

	Neuschwanstein	Příbram
α_R (apparent right ascension)	$190.59^\circ \pm 0.03^\circ$	$190.121^\circ \pm 0.009^\circ$
δ_R (apparent declination)	$21.98^\circ \pm 0.03^\circ$	$20.425^\circ \pm 0.001^\circ$
v_∞ (initial velocity)	$20.95 \pm 0.04 \text{ km s}^{-1}$	$20.886 \pm 0.005 \text{ km s}^{-1}$
α_G (geocentric right ascension)	$192.33^\circ \pm 0.03^\circ$	$192.338^\circ \pm 0.010^\circ$
δ_G (geocentric declination)	$19.54^\circ \pm 0.04^\circ$	$17.467^\circ \pm 0.002^\circ$
v_G (geocentric velocity)	$17.51 \pm 0.05 \text{ km s}^{-1}$	$17.431 \pm 0.006 \text{ km s}^{-1}$
v_H (heliocentric velocity)	$37.46 \pm 0.04 \text{ km s}^{-1}$	$37.451 \pm 0.005 \text{ km s}^{-1}$
a (semimajor axis)	$2.40 \pm 0.02 \text{ AU}$	$2.401 \pm 0.002 \text{ AU}$
e (eccentricity)	0.670 ± 0.002	0.6711 ± 0.0003
q (perihelion distance)	$0.7929 \pm 0.0004 \text{ AU}$	$0.78951 \pm 0.00006 \text{ AU}$
Q (aphelion distance)	$4.01 \pm 0.03 \text{ AU}$	$4.012 \pm 0.005 \text{ AU}$
ω (argument of perihelion)	$241.20^\circ \pm 0.06^\circ$	$241.750^\circ \pm 0.013^\circ$
Ω (longitude of ascending node)	$16.82664^\circ \pm 0.00001^\circ$	$17.79147^\circ \pm 0.00001^\circ$
i (inclination)	$11.41^\circ \pm 0.03^\circ$	$10.482^\circ \pm 0.004^\circ$

either in catalogues of known near-Earth objects, or among meteorites and fireballs observed in the past. Though none of the orbits of previous fireballs matches the orbits of Příbram and Neuschwanstein exactly, we find several fireballs that could be tentatively associated with the pair. These are fireballs EN080486 ($D' = 0.033$), PN38856 ($D' = 0.065$), PN39590 ($D' = 0.057$) and MORP300 ($D' = 0.084$), as well as the Glanerbrug L-LL chondrite ($D' = 0.1$) which fell on 7 April 1990 in the Netherlands, and whose orbit was recovered from visual observations¹⁶. Judging from the atmospheric behaviour of these fireballs and from the chemistry of Glanerbrug, we see large structural variety, suggesting that objects in the 'Příbram stream' vary greatly in their chemical composition, structure and ages.

Among the Earth-approaching asteroids, two objects—4486 Mithra ($D' = 0.078$) and 1998SJ₇₀ ($D' = 0.104$)—have previously

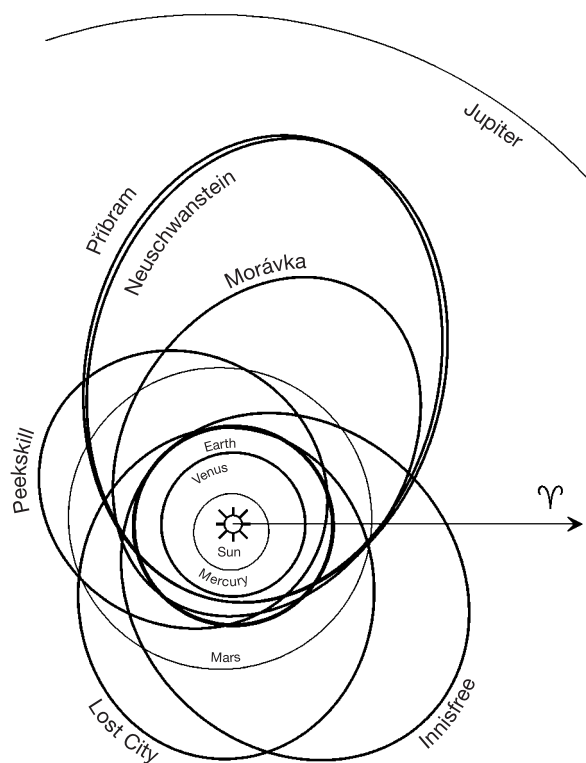


Figure 3 Comparison of meteorite orbits. The figure shows the orbit of the Neuschwanstein meteorite, along with the three meteorites with photographically determined orbits (Příbram¹, Lost City²² and Innisfree²¹) and two meteorites with orbits determined from video records (Peekskill²³ and Morávka⁶). All orbits are projected onto the ecliptic plane. γ indicates the vernal equinox.

been identified as possible associates of Příbram^{11,12}. In summer 2000, 4486 Mithra was well placed for observations by radar from Arecibo and Goldstone. The object was found to be in unusual spin state, double-lobed, and more severely bifurcated than any other near-Earth asteroid targeted previously¹⁷. Such objects are typically believed to be 'rubble piles', loosely bound by gravitation. During close encounters with planets, these objects would be exposed to tidal forces that could lead to their disruption or to the detachment of significant amounts of near-surface rubble that would then be found in orbits similar to that of the parent. This model could possibly resolve the paradox of young streams containing different classes of objects of long cosmic-ray exposure.

Further discussions of the origin of the 'Příbram stream' must await the final results of the laboratory analyses of Neuschwanstein; more data could be gained from clear identification of other members of this stream (either meteorites or asteroids) in the future. The Neuschwanstein and Příbram meteorites bring new information to the discussions of the mechanisms of meteorite delivery to the Earth, and demonstrate the importance of long-term fireball observing programmes. □

Received 23 January; accepted 28 March 2003; doi:10.1038/nature01592.

1. Ceplecha, Z. Multiple fall of Příbram meteorites photographed. *Bull. Astron. Inst. Czech.* **12**, 21–47 (1961).
2. Bischoff, A. & Zipfel, J. Mineralogy of the Neuschwanstein (EL6) chondrite—first results. *Lunar Planet. Sci. Conf. XXXIV*, abstr. 1212 (2003).
3. Ceplecha, Z. Geometric, dynamic, orbital and photometric data on meteoroids from photographic fireball networks. *Bull. Astron. Inst. Czech.* **38**, 222–234 (1987).
4. Borovička, J., Spurný, P. & Keckliková, J. A new positional astrometric method for all-sky cameras. *Astron. Astrophys. Suppl. Ser.* **112**, 173–178 (1995).
5. Ceplecha, Z., Spurný, P., Borovička, J. & Keckliková, J. Atmospheric fragmentation of meteoroids. *Astron. Astrophys.* **279**, 615–626 (1993).
6. Borovička, J., Spurný, P. & Ceplecha, Z. The Morávka meteorite fall: Fireball trajectory, orbit and fragmentation from video records. *Meteorit. Planet. Sci. Suppl.* **36**, A25–A26 (2001).
7. Drummond, J. D. A test of comet and meteor associations. *Icarus* **45**, 545–553 (1981).
8. Halliday, I., Griffin, A. A. & Blackwell, A. T. in *Highlights of Astronomy* Vol. 6 (ed. West, R. M.) 399–404 (D. Reidel, Dordrecht, 1983).
9. Halliday, I. Detection of a meteorite "stream": Observations of a second meteorite fall from the orbit of the Innisfree chondrite. *Icarus* **69**, 550–556 (1987).
10. Halliday, I., Blackwell, A. T. & Griffin, A. A. Evidence for the existence of groups of meteorite-producing asteroidal fragments. *Meteorit. Planet. Sci.* **25**, 93–99 (1990).
11. Drummond, J. D. Earth-approaching asteroid streams. *Icarus* **89**, 14–25 (1991).
12. Drummond, J. D. The D discriminant and near-Earth asteroid streams. *Icarus* **146**, 453–475 (2000).
13. Dodd, R. T., Wolf, S. F. & Lipschutz, M. E. A H chondrite stream: Identification and confirmation. *J. Geophys. Res.* **98**, 15105–15118 (1993).
14. Treiman, A. H. An improbable concentration of basaltic meteorite falls (HED and mesosiderite) in the mid-20th century. *Meteorit. Planet. Sci.* **28**, 246–252 (1993).
15. Stauffer, H. & Urey, H. C. Multiple fall of Příbram meteorites photographed III. Rare gas isotopes in the Velka stone meteorite. *Bull. Astron. Inst. Czech.* **13**, 106–109 (1962).
16. Jenniskens, P. *et al.* Orbits of meteorite producing fireballs. The Glanerbrug—a case study. *Astron. Astrophys.* **255**, 373–376 (1992).
17. Ostro, S. J. *et al.* Radar observations of asteroid 4486 Mithra. *DPS meeting 32* (American Astronomical Society, Pasadena, California, 2000).
18. Ceplecha, Z. Fireballs photographed in central Europe. *Bull. Astron. Inst. Czech.* **28**, 328–340 (1977).
19. Oberst, J. *et al.* The "European Fireball Network": Current status and future prospects. *Meteorit. Planet. Sci.* **33**, 49–56 (1998).
20. McCrosky, R. E. & Boeschstein, H. The Prairie Meteorite Network. *Smithson. Astrophys. Obs. Spec. Rep.* **173**, 1–23 (1965).
21. Halliday, I., Blackwell, A. T. & Griffin, A. A. The Innisfree meteorite and the Canadian camera network. *J. R. Astron. Soc. Can.* **72**, 15–39 (1978).
22. McCrosky, R. E., Posen, A., Schwartz, G. & Shao, C.-Y. Lost City meteorite—Its recovery and a comparison with other fireballs. *J. Geophys. Res.* **76**, 4090–4108 (1971).
23. Brown, P. *et al.* The orbit and atmospheric trajectory of the Peekskill meteorite from video records. *Nature* **367**, 624–626 (1994).

Acknowledgements We thank the local operators at European Fireball Network (EN) stations for their service at the time of the Neuschwanstein fall, and the finder of the meteorite for the positional information. We also thank J. Keckliková for discussions and contributions to analysis software. The German part of the EN is operated by the German Aerospace Center (DLR), Berlin, the Czech part of the EN is operated by the Astronomical Institute of the Academy of Sciences of the Czech Republic, Ondřejov, and the Austrian station is maintained by the Gahberg Observatory.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.S. (spurny@asu.cas.cz).

Controlled vesicle deformation and lysis by single oscillating bubbles

Philippe Marmottant & Sascha Hilgenfeldt

Faculty of Applied Physics, University of Twente, PO Box 217, 7500AE Enschede, The Netherlands

The ability of collapsing (cavitating) bubbles to focus and concentrate energy, forces and stresses is at the root of phenomena such as cavitation damage, sonochemistry or sonoluminescence^{1,2}. In a biomedical context, ultrasound-driven microbubbles have been used to enhance contrast in ultrasonic images³. The observation of bubble-enhanced sonoporation^{4–6}—acoustically induced rupture of membranes—has also opened up intriguing possibilities for the therapeutic application of sonoporation as an alternative to cell-wall permeation techniques such as electroporation⁷ and particle guns⁸. However, these pioneering experiments have not been able to pinpoint the mechanism by which the violently collapsing bubble opens pores or larger holes in membranes. Here we present an experiment in which gentle (linear) bubble oscillations are sufficient to achieve rupture of lipid membranes. In this regime, the bubble dynamics and the ensuing sonoporation can be accurately controlled. The use of microbubbles as focusing agents makes acoustics on the micrometre scale (microacoustics) a viable tool, with possible applications in cell manipulation and cell-wall permeation as well as in microfluidic devices.

Most experiments assessing sonoporation have used strong ultrasonic fields giving rise to a multitude of simultaneous phenomena. The micrographs of ref. 4 suggest that the formation of microjets may be involved, a process that is very difficult to control and model^{9–11}. Other candidates for damaging agents during bubble collapse involve shockwaves emitted from the bubble¹² and/or subsequent heating of the cell wall¹³. All of these processes are significant only when the bubble undergoes a fast, inertial collapse. By contrast, some experiments¹⁴ demonstrated cell transfection with moderate, off-resonance driving which should have resulted in much weaker bubble oscillations. Our experiments use single microbubbles fixed on a substrate, and use far smaller ultrasound driving amplitudes, replacing the nonlinear dynamics of the inertial collapse with gentle, linear oscillations. Linear oscillations can be sufficient to rupture single cells, because the bubbles' response concentrates ultrasonic energy on the microscale, whereas it could conventionally only be focused to about an ultrasonic wavelength (a few centimetres or millimetres).

In our set-up (Fig. 1a), bubbles of about 10–100 µm radius are attached by capillary forces to the walls of a quartz cuvette (10 mm × 10 mm × 40 mm), after they were generated by syringe injection of air. A piezoelectric transducer provides a standing-wave ultrasound field inside the cuvette, which is filled with a suspension of either cells or lipid vesicles. We chose giant unilamellar lipid vesicles ('artificial cells' consisting only of a lipid bilayer membrane) as our primary object of study because of their well-defined mechanical properties^{15,16}. Using electroformation¹⁷, we obtained 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) vesicles with radii of 10–100 µm. We use fluorescence marking of either the lipid itself or the interior of the vesicles (when grown in a fluorescent dye solution and then transferred to the experiment). A high-speed camera (≤2,000 frames per second) connected to an inverted microscope records phase contrast or fluorescence images of the vesicles' reaction to the bubble oscillation. When not using fluorescence, the contrast between bubbles and solution is enhanced by growing the vesicles in a sucrose solution and transferring them to a

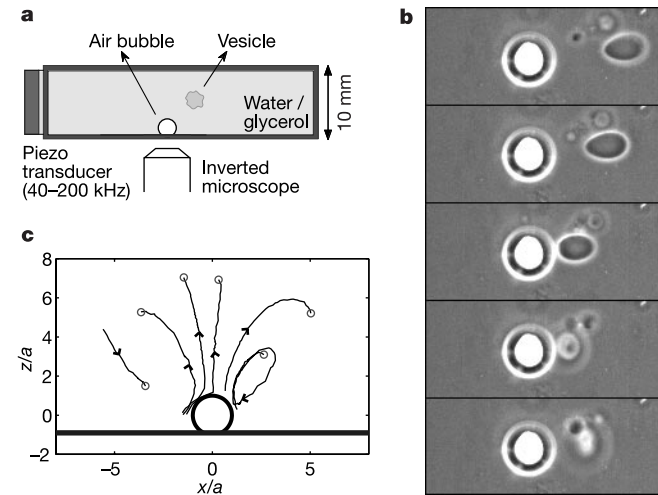


Figure 1 Vesicle motion near an oscillating bubble. **a**, Layout of the experiment. The piezoelectric transducer generates a standing ultrasound field. **b**, High-speed time series (bottom view through inverted microscope) of vesicle motion (interframe time 10 ms). The bright object is a bubble of radius $a \approx 15 \mu\text{m}$ at the cuvette wall, whose oscillation amplitude is too small to be seen here. The dark object on the right is a lipid vesicle, whose shape far from the bubble would be spherical. Here it is severely deformed as it approaches the bubble, collides with it, and is then expelled away from the observer (upwards in the cuvette), blurring as it leaves the focal plane (see Supplementary movie 1). **c**, Experimentally observed trajectories of vesicles in a side view (z is the axis perpendicular to the cuvette wall to which the bubble is attached).

glucose solution of equal osmolality, but different index of refraction.

Figure 1b shows a phase contrast image sequence of a vesicle (radius $R \approx 10 \mu\text{m}$) next to a bubble of radius $a \approx 15 \mu\text{m}$ (bright object) attached to the cuvette wall. (See also Supplementary movie 1.) The ultrasound driving at frequency $f = 180 \text{ kHz}$ is near-resonant for the bubble, but its amplitude is low: the amplitude of bubble oscillation, measured by comparing bubble sizes with and without driving ultrasound, is only εa , with $\varepsilon \leq 0.05$. From linearized Rayleigh–Plesset bubble dynamics^{18,19}, we infer a driving pressure amplitude of about 0.1 bar. Nevertheless, the vesicle reacts with vigorous motion around the bubble (interframe time in Fig. 1b is 10 ms), being alternately attracted and repelled. Observations with the cuvette tilted by 90° (now viewing tangentially along the cuvette wall) show that the motion occurs in a plane perpendicular to the wall (Fig. 1c). The vesicles follow loop-like trajectories, approaching the bubble near the wall and then being repelled radially outwards and upwards, away from the wall. The period of this cyclic motion is typically 0.1–1 s, much longer than the ultrasound period. After a few minutes, many vesicles tend to accumulate around the bubble and perform the same motion. They are also deformed in a characteristic fashion: spherical vesicles become prolate when approaching the bubble (see Fig. 1b) and oblate when receding.

When the amplitude of driving is doubled, the speed of the vesicle increases fourfold. This and the observed timescales of the motion suggest that it is a second-order effect of the nonlinearity in the Navier–Stokes equations (rather than an effect of nonlinear bubble dynamics). Modelling can be attempted using the well-established theory of acoustic streaming²⁰, where a second-order steady flow emerges from an oscillatory primary flow. Streaming flow from air pockets was suspected in the aggregation and damage of blood platelets and cells around microfilter pores^{21,22}. However, the observed recirculation patterns failed to match the then-existing theory²³ even qualitatively. Elder²³ explained the streaming of aluminium particles around millimetre-sized bubbles. Lighthill²⁰

remarks that in all such engineering applications, the Reynolds number of the steady flow is large (‘Stuart streaming’). By contrast, the considerably smaller (micrometre) scales of the present experiments (and of those in refs 21 and 22) lead to a different regime of streaming motion. The bubble-streaming Reynolds number²⁴ is:

$$\text{Re}_{\text{bs}} = \varepsilon^2 \left(\frac{\omega a^2}{\nu} \right)^{1/2} \quad (1)$$

where $\omega = 2\pi f$ is the angular frequency and $\nu \approx 1.0 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}$ is the kinematic viscosity of the aqueous solution. With the values given above, we obtain $\text{Re}_{\text{bs}} \approx 0.04$, much less than in previous microstreaming studies of haemolysis²⁵. The secondary flow is therefore Rayleigh–Nyborg–Westervelt (RNW) streaming²⁰, which is considerably easier to treat theoretically, as it is an example of Stokes flow.

In our case, the primary oscillation is due to a superposition of bubble volume oscillations and bubble translational oscillations: as the bubble volume changes periodically while its bottom remains fixed to the wall, the centre of mass executes up-and-down motions with amplitude $\varepsilon' \approx \varepsilon$ (Fig. 2a). Longuet-Higgins²⁶ has calculated the steady RNW streaming flow of such a bubble motion in bulk liquid, resulting in the second-order stream function:

$$\psi_2^s = \frac{\varepsilon^2}{4} a^3 \omega \sin(\Delta\phi) \left(-2\frac{r}{a} + \frac{a}{r} + \frac{a^4}{r^4} \right) \sin^2 \theta \quad (2)$$

Here, $\Delta\phi$ is the phase shift between volume and translational oscillations, and r the distance to the bubble centre, while θ is the angle with the axis of translation (that is, gravity). The flow is thus dominated by a finite number of Stokes singularities, in particular a stokeslet ($\psi \propto r \sin^2 \theta$) and a potential dipole ($\psi \propto r^{-1} \sin^2 \theta$).

It is now easy to introduce the presence of the cuvette wall using the method of images in Stokes flow^{27,28}, where a number of image

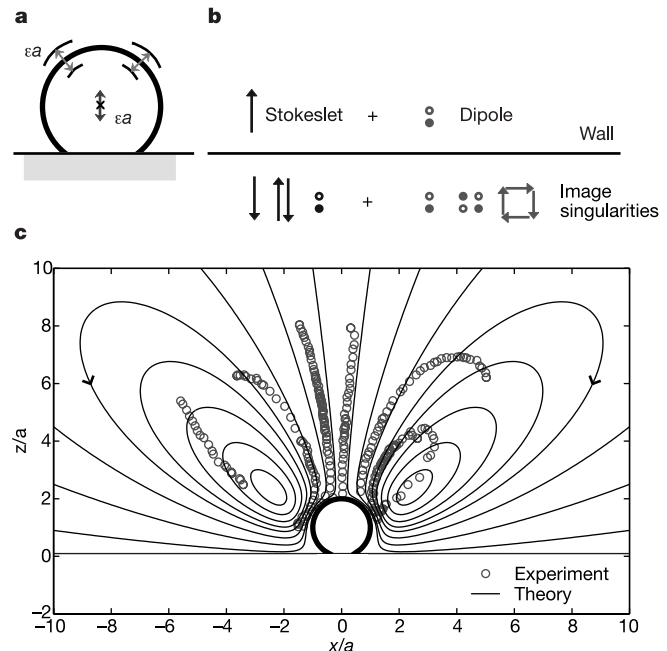


Figure 2 Modelling of the streaming flow. **a**, A bubble at a wall undergoes both volume and translational oscillations. **b**, The streaming flow is composed of leading-order stokeslet and dipole singularities. They require three image singularities each to compute the complete flow field in the presence of the wall. **c**, Streamlines of the computed flow field (for $z > 0$), superimposed on a number of measured vesicle trajectories, scaled properly. Deviations between experiment and theory show the influence of weak streaming around the vesicles.

singularities are placed at the opposite side of the wall to make all velocity components at the wall vanish (Fig. 2b). The Stokes flow fields of these singularities can be simply added up to yield an analytic solution for the flow field. The leading-order far-field term of the complete solution is a dipole-like streaming:

$$\psi_2^{\text{wall}} = -3\varepsilon^2 a^3 \omega \sin(\Delta\phi) \frac{a}{r} \cos^2 \theta \sin^2 \theta + O(\varepsilon^2 r^{-2}) \quad (3)$$

The streamlines of the complete solution are plotted in Fig. 2c. They are very similar to the vesicle trajectories and illustrate that the vesicles essentially follow the flow as passive tracers. Indeed, as the vesicles (1) have a lipid bilayer in a liquid state, and (2) have a much smaller compressibility modulus than bubbles, the streaming around them should be weaker by up to a factor of $\delta = (\nu/\omega a^2)^{1/2}$ compared to the streaming around volume-oscillating bubbles^{24,26,29}. For the numbers given above, $\delta \approx 0.06$. The liquid character of the lipid bilayer also leads to tank-treading motion, which is readily observed in our experiments with fluorescence-marked lipid membranes.

The streaming velocities inferred from the complete version of equation (3) are in quantitative agreement with those observed (with a maximum velocity close to the bubble of $u \approx \varepsilon^2 a \omega \approx 1 \text{ mm s}^{-1}$). Some vesicles (particularly large ones) seek out the stagnation line of the flow field (at a distance of about $2a$ from the bubble), where their centre of mass stays stationary while the membrane is undergoing tank-treading and deformation. This situation is particularly well-suited for a controlled change of the forcing parameters to control the degree of deformation of the vesicle. The deformation is determined by the shear rate of the streaming flow field (computed from the rate-of-strain tensor $G_{ij} = (\partial_i u_j + \partial_j u_i)/2$) and the vesicle membrane elasticity have to be accounted for¹⁶.

In the experiments presented so far, the maximum rate of strain as calculated from the theory is $G_{\text{max}} \approx 10^3 \text{ s}^{-1}$. This leads to a maximum tension in the membrane of $\tau_{\text{max}} \approx \eta G_{\text{max}} R \approx 10^{-4} \text{ N m}^{-1}$, where $\eta \approx 10^{-3} \text{ kg m}^{-1} \text{ s}^{-1}$ is the dynamic viscosity of the aqueous solution. This tension is sufficient to iron out all the thermal

fluctuations of the membrane area, but will only expand the actual membrane area by a fraction $\Delta A/A \approx 4 \times 10^{-4}$, assuming an area expansion modulus of DOPC of about 0.24 N m^{-1} (ref. 16) and using the theory of Seifert³⁰. To achieve rupture of the membrane (or at least opening of pores), $\Delta A/A$ has to be as large as 0.03 (ref. 16). The present theory therefore predicts that a 100-fold increase in ηGR will lead to vesicle rupture.

We tested this prediction by performing the experiment in a water/glycerol mixture with $\eta = 20 \eta_{\text{water}}$ and using slightly larger driving amplitudes and vesicle radii. Figure 3 shows two fluorescence imaging series in which rupture can indeed be observed. In Fig. 3a, the vesicle membrane is fluorescence-marked; as the vesicle approaches the bubble, it develops a conical tip in the region of maximum shear rate and then ruptures, with fluorescent debris surrounding the bubble. In Fig. 3b, the liquid inside the vesicle was fluorescence-marked. This vesicle was stationary (tank-treading) close to the bubble when the driving amplitude was increased. The vesicle again shows tip formation and then undergoes complete lysis over an easily controllable timescale of about 1 s. (Also see Supplementary movie 2.) Less dramatic cases of dye leakage have also been observed, which is promising in view of the desired controlled exchange of interior and exterior liquid in drug and DNA delivery applications.

The use of microbubbles as focusing agents makes acoustics a viable tool on the micrometre scale ('microacoustics'). The acoustic streaming flow exerts large enough shear forces on vesicles in the flow to manipulate, deform, and rupture them. With the known parameters from the lipid vesicle experiments, we can now assess the properties of cells, for example, their rupture strain. Preliminary experiments have shown that deformation and lysis of whole HeLa cells is possible in this set-up. Cell viability tests through staining are currently underway. The detailed control over the linear bubble oscillations allows for a quantitative fine-tuning and optimization of this method for cell manipulation, cell wall permeation, or cell lysis. The microstreaming generated by the bubbles offers other interesting perspectives: 'acoustical tweezers' or even microfluidics devices based on this principle become feasible (P. M. and S. H., unpublished work). □

Received 27 December 2002; accepted 28 March 2003; doi:10.1038/nature01613.

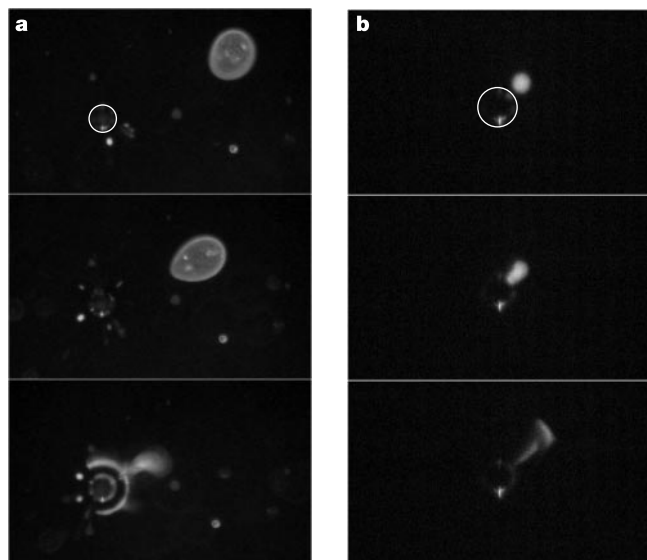


Figure 3 Vesicle rupture. **a**, A vesicle ($R \approx 40 \mu\text{m}$) with fluorescence-marked lipid membrane approaches a bubble (marked by the white circle, $a \approx 20 \mu\text{m}$). The vesicle deforms and fragments. Interframe time is 1 s, ultrasound frequency is 40 kHz. **b**, A tank-treading vesicle ($R \approx 15 \mu\text{m}$) at a stationary position near a bubble (white circle, $a \approx 20 \mu\text{m}$). The driving amplitude was increased at about the time of the first frame. The liquid inside the vesicle is fluorescent here, and disperses after vesicle lysis (last frame). Interframe time of 0.4 s, ultrasound frequency $f = 180 \text{ kHz}$ (see Supplementary movie 2).

- Suslick, K. S. Sonochemistry. *Science* **247**, 1439–1445 (1990).
- Hilgenfeldt, S., Grossmann, S. & Lohse, D. A simple explanation of light emission in sonoluminescence. *Nature* **398**, 402–405 (1999).
- Chang, D. C., Chassey, B. M., Saunders, J. A. & Sowers, A. E. (eds) *Advances in Echo Imaging Using Contrast Enhancement* (Kluwer Academic, Dordrecht, 1993).
- Tachibana, K., Uchida, T., Ogawa, K., Yamashita, N. & Tamura, K. Induction of cell membrane porosity by ultrasound. *Lancet* **353**, 1409 (1999).
- Miller, D. L. & Qudus, J. Sonoporation of monolayer cells by diagnostic ultrasound activation of contrast-agent gas bodies. *Ultrasound Med. Biol.* **26**, 661–667 (2000).
- Ward, M., Wu, J. & Chiu, J.-F. Experimental study of the effects of Optison® concentration on sonoporation *in vitro*. *Ultrasound Med. Biol.* **26**, 1169–1175 (2000).
- Chang, D. C., Chassey, B. M., Saunders, J. A. & Sowers, A. E. (eds) *Guide to Electroporation and Electrofusion* (Academic, New York, 1992).
- Seki, M., Komeda, Y., Iida, A., Yamada, Y. & Morikawa, H. Transient expression of beta-glucuronidase in *Arabidopsis thaliana* leaves and roots and *Brassica napus* stems using a pneumatic particle gun. *Plant Mol. Biol.* **17**, 259–263 (1991).
- Prosperetti, A. A new mechanism for sonoluminescence. *J. Acoust. Soc. Am.* **101**, 2003–2007 (1997).
- Popinet, S. & Zaleski, S. A front-tracking algorithm for accurate representation of surface tension. *Int. J. Num. Meth. Fluids* **30**, 775–793 (1999).
- Brujan, E.-A., Nahen, K., Schmidt, P. & Vogel, A. Dynamics of laser-induced cavitation bubbles near an elastic boundary. *J. Fluid Mech.* **433**, 251–281 (2001).
- Wang, Z. Q., Pecha, R., Gompf, B. & Eisenmenger, W. Single bubble sonoluminescence: Investigations of the emitted pressure wave with a fiber optic probe hydrophone. *Phys. Rev. E* **59**, 1777–1780 (1999).
- Hilgenfeldt, S., Lohse, D. & Zomack, M. Sound scattering and localized heat deposition of pulse-driven microbubbles. *J. Acoust. Soc. Am.* **107**, 3530–3539 (2000).
- Greenleaf, W. J., Bolander, M. E., Sarkar, G., Goldring, M. B. & Greenleaf, J. F. Artificial cavitation nuclei significantly enhance acoustically induced cell transfection. *Ultrasound Med. Biol.* **24**, 587–595 (1998).
- Rawicz, W., Olbrich, K., McIntosh, T., Needham, D. & Evans, E. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophys. J.* **79**, 328–339 (2000).
- Boal, D. *Mechanics of the Cell* (Cambridge Univ. Press, Cambridge, UK, 2002).
- Dimitrov, D. S. & Angelova, M. I. Lipid swelling and liposome formation on solid surfaces. *Prog. Colloid Polym. Sci.* **73**, 48–56 (1987).

18. Plesset, M. S. & Prosperetti, A. Bubble dynamics and cavitation. *Annu. Rev. Fluid Mech.* **9**, 145–185 (1977).
19. Leighton, T. G. *The Acoustic Bubble* (Academic, London, 1994).
20. Lighthill, J. Acoustic streaming. *J. Sound Vib.* **61**, 391–418 (1978).
21. Miller, D. L., Nyborg, W. L. & Whitcomb, C. C. Platelet aggregation induced by ultrasound under specialized conditions. *Science* **205**, 505–507 (1979).
22. Miller, D. L. Particle gathering and microstreaming near ultrasonically activated micropores. *J. Acoust. Soc. Am.* **84**, 1378–1387 (1988).
23. Elder, S. A. Cavitation microstreaming. *J. Acoust. Soc. Am.* **31**, 54–64 (1958).
24. Davidson, B. J. & Riley, N. Cavitation microstreaming. *J. Sound Vib.* **15**, 217–233 (1971).
25. Rooney, J. A. Hemolysis near an ultrasonically pulsating gas bubble. *Science* **169**, 869–871 (1970).
26. Longuet-Higgins, M. S. Viscous streaming from an oscillating spherical bubble. *Proc. R. Soc. Lond. A* **454**, 725–742 (1998).
27. Blake, J. R. & Chwang, A. T. Fundamental singularities of viscous flow. *J. Eng. Math.* **8**, 23–29 (1974).
28. Pozrikidis, C. *Boundary Integral and Singularity Methods for Linearized Viscous Flow* (Cambridge Univ. Press, Cambridge, UK, 1992).
29. Riley, N. On a sphere oscillating in a viscous fluid. *Q. J. Mech. Appl. Math.* **19**, 461–472 (1966).
30. Seifert, U. Fluid membranes in hydrodynamic flow fields: Formalism and an application to fluctuating quasispherical vesicles in shear flow. *Eur. Phys. J. B* **8**, 405–415 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Lohse for his support and insight, and A. van den Berg and H. Gardeniers for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P. M. (p.marmottant@tn.utwente.nl).

Spreading of nanofluids on solids

Darsh T. Wasan & Alex D. Nikolov

Department of Chemical and Environmental Engineering, Illinois Institute of Technology, Chicago, Illinois 60616, USA

Suspensions of nanometre-sized particles (nanofluids) are used in a variety of technological contexts. For example, their spreading and adhesion behaviour on solid surfaces can yield materials with desirable structural and optical properties¹. Similarly, the spreading behaviour of nanofluids containing surfactant micelles has implications for soil remediation, oily soil removal, lubrication and enhanced oil recovery. But the well-established concepts of spreading and adhesion of simple liquids do not apply to nanofluids^{2–7}. Theoretical investigations have suggested that a solid-like ordering of suspended spheres will occur in the confined three-phase contact region at the edge of the spreading fluid, becoming more disordered and fluid-like towards the bulk phase^{8,9}. Calculations have also suggested that the pressure arising from such colloidal ordering in the confined region will enhance the spreading behaviour of nanofluids^{10,11}. Here we use video microscopy to demonstrate both the two-dimensional crystal-like ordering of charged nanometre-sized polystyrene spheres in water, and the enhanced spreading dynamics of a micellar fluid, at the three-phase contact region. Our findings suggest a new mechanism for oily soil removal—detergency.

When a gas bubble or oil/liquid drop dispersed in an aqueous nanofluid approaches a smooth, horizontal hydrophilic solid surface, there is a microscopic transition between the liquid film and the meniscus—and the nanofluid film can change its thickness in steps¹¹. The transition region between the liquid film on a solid surface and the bulk meniscus has a wedge-like profile (Fig. 1), and its shape is determined by forces arising from the ordering of particles or supermolecules, such as surfactant micelles. Structural transitions have been observed in colloidal fluids confined to very thin wedges^{12–15}.

The aim of this study is to reveal the effects of the particle

structure formation and the structural disjoining pressure on the spreading of colloidal fluids on solid surfaces. Spreading is generally described in terms of the spreading coefficient S , given by $S = \sigma_{sg} - \sigma_{sl} - \sigma_{lg}$, where σ is the respective interfacial tension existing between the solid–gas (σ_{sg}), solid–liquid (σ_{sl}) and liquid–gas (σ_{lg}) phases. The de Gennes theory relates the value of S to the disjoining pressure of a wetting film ($\Pi(h)$, where h is the thickness of the film) as^{2,16}:

$$S = \Pi(h)h + \int_h^\infty \Pi(h) dh \quad (1)$$

We directly observed the particle-structuring phenomenon in the liquid film–meniscus region by using reflected-light digital video microscopy. The wedge film was formed by blowing a small air bubble (diameter 200 μm) against an optically smooth glass plate in a suspension of monodisperse, charged latex spheres in deionized surfactant-free water. The volume fraction of spheres in the fluid

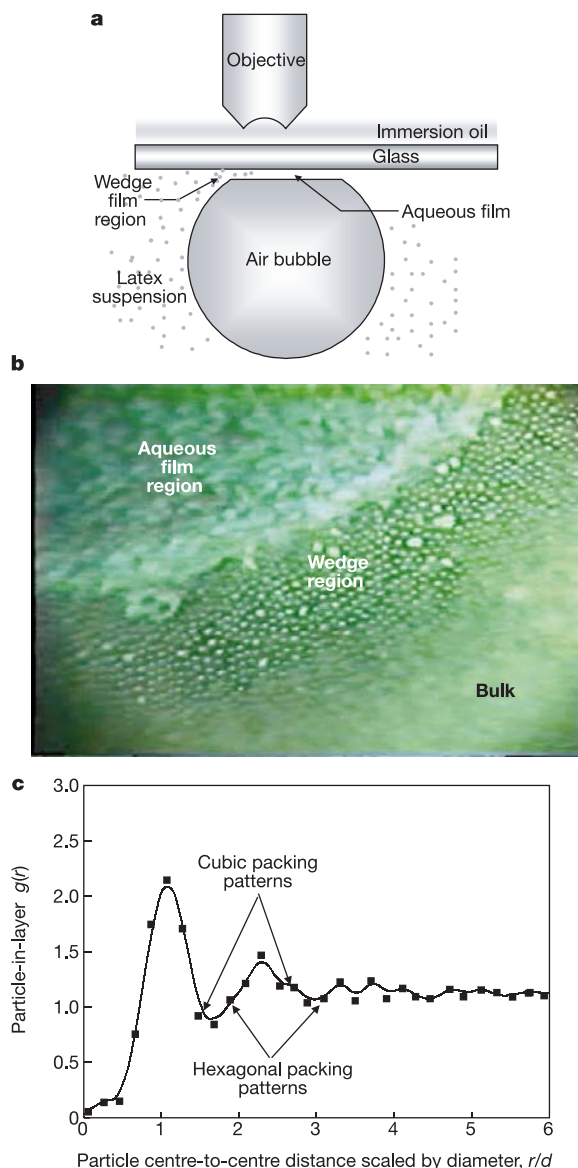


Figure 1 Particle structuring in a wedge film. **a**, Diagram of experimental set-up. **b**, Particle structuring in a wedge film. Latex particles had diameter 1 μm , charge $0.8 \mu\text{C cm}^{-2}$, and occupied 7 vol.%. **c**, In-layer particle structure inside the wedge film. Mean particle-to-particle distance was $1.2d$, effective particle volume was 44 vol.%.

was 7 vol.%, and they had a surface charge of $0.8 \mu\text{C cm}^{-2}$, a diameter of $1.01 \mu\text{m}$, and a density of 1.05 g cm^{-3} . Figure 1a shows a diagram of the experimental set-up. Special precautions were taken to eliminate light coming from the air-bubble/liquid surface as well as light reflected from the glass/air surface and heating effects. The optical signal was digitized for analysis. Figure 1b shows a photomicrograph taken in reflected light depicting the formation of a two-dimensional (2D), in-layer, colloidal crystal structure in the three-phase contact region (that is, in the wedge film). The latex particles tend to form a 2D colloid crystal structure at a thickness of the wedge film equal to twice the particle diameter (that is, about $2 \mu\text{m}$). However, the particle in-layer structure changes to a disordered structure when the film thickness exceeds three particle diameters (see Supplementary Information movie 1). The co-existence of this ordered-disordered structure depicted in Fig. 1b is reproducible, and was observed for several hours.

Figure 1c shows the particle in-layer radial distribution function, $g(r)$, that we obtained from the image analysis of over 2,000 particles. This figure shows the co-existence of both the 2D hexagonal and cubic packing domains in the wedge film.

We also used computer-enhanced video microscopy to monitor the position of the particle oscillations in the three-phase contact region (Supplementary Information movie 1). The brownian motion of the colloidal latex particles along the x and y axes were measured in more than 600 successive time intervals of 0.033 s . The period of oscillation was about 2 Hz . Plots of the probability density distribution function for the disordered and ordered particle structures are shown in Fig. 2a and b, respectively. The results

show that the particle displacements due to brownian motion follow a gaussian distribution. However, we note that the particle mean-square displacement (standard deviation) for the ordered and disordered structures is very different. The value of the mean-square displacement for the 2D colloidal crystal-like structure is $0.17d$ (where d is particle diameter) compared to $0.9d$ for the random-packed structure. The mean interparticle distance is $1.2d$.

We also estimated the energy of interaction between the particles in the three-phase contact region using the Boltzman equation. The energy for particle crystallization (that is, changing particles from a random-packed structure to a 2D colloid crystal-like structure) was calculated to be only $1.8kT$ (where k is the Boltzmann constant, and T is temperature). The particle-particle mean interaction potential becomes oscillatory owing to the particle crystallization in the wedge film, so the film disjoining pressure (Π) versus film thickness (h) has an oscillatory decay.

In Fig. 3a we plot disjoining pressure, obtained by using an analytical expression based on statistical mechanics¹⁰. The pressure is high (nearly $50,000 \text{ Pa}$) near the vertex, because the hard spheres are 'pushed' into the region by the spheres farther out. The pressure oscillates as 1, 2, 3 and 4 layers of hard spheres can be accommodated between the wedge confinement, and ultimately becomes the bulk hard-sphere pressure. The magnitude of this pressure depends on the effective particle volume fraction and particle size. We confirmed these integral equation results by comparison with Monte Carlo simulation results⁸. Computer simulations have shown that there is a tendency for the spheres near the vertex to have a solid-like structure, and for those further from the vertex to have a fluid-like structure^{8,9}. These theoretical results confirm our experimental observations on the particle structure formation in the wedge film (Fig. 1).

The spreading coefficient, S , was estimated as a function of the number of particle layers in the wedge film using equation (1) and the analytical expressions derived from the Ornstein-Zernike

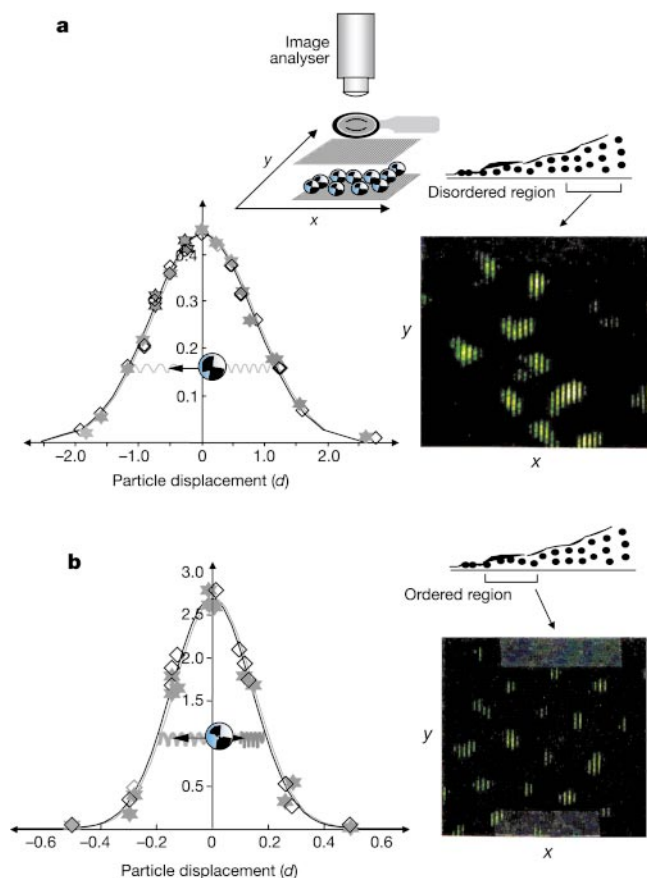


Figure 2 Distribution of particle displacement inside the wedge film. **a**, Inside the 2D random packing region; **b**, inside the 2D 'crystal' region. In graphs, stars and diamonds show respectively x and y displacements. Video prints show instantaneous particle positions.

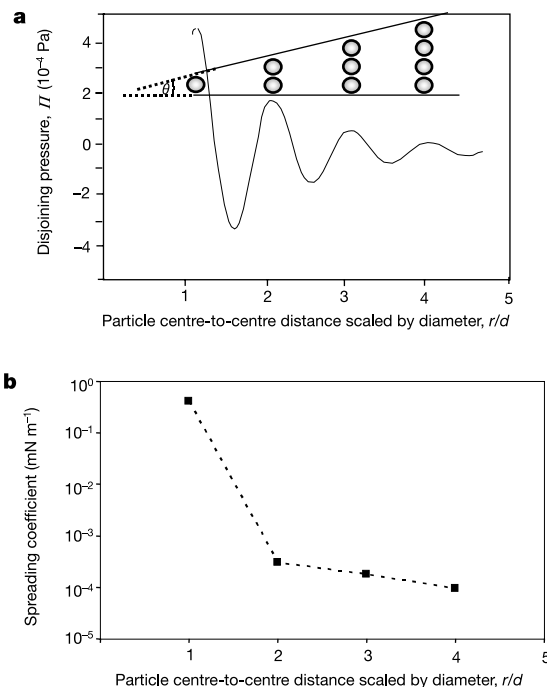


Figure 3 Pressure profile and spreading coefficient as a function of film thickness.

a, Pressure profile on the wedge walls as a function of film thickness r scaled by particle diameter d . Particle volume fraction, 0.4; particle diameter $d = 8 \text{ nm}$.

b, Spreading coefficient arising owing to particle disjoining pressure. Nanofluid properties as in **a**.

equation for the structural disjoining pressure in the film¹⁰ (Fig. 3a). Our calculations show that the spreading coefficient increases with a decrease in film thickness, that is, with the decrease in the number of particle layers inside the film (Fig. 3b). Also, we notice a significant change in the slope of the curve at a thickness of the wedge film equal to twice the particle diameter. Our Monte Carlo simulations showed that at this film thickness the particle in-layer structure changes to an ordered structure¹¹. These results indicate that the in-layer particle structuring can enhance the spreading of nanofluids on solids.

Next, we studied the spreading dynamics of micellar fluids and role of the structural component of the disjoining pressure arising from the nanoparticle (micelle) ordering inside the film-meniscus region (that is, the wedge film) in the oily soil removal process. We studied the oil/solid interactions in the presence of an aqueous micellar solution, using differential and common three-phase contact angle interferometry^{17–22}. Figure 4a shows the interference patterns produced by an oil droplet (of volume 10 μl) on a glass/air surface observed by reflected light using a differential microscope (Supplementary Information movie 2). The interference patterns seen in this photomicrograph as parallel, regular, continuous lines depict the topology of the wedge film region. The distance between the patterns and their order of interference can be used to calculate the local radii of curvature. These radii, in conjunction with the surface tension data, allow us to calculate the capillary pressure. We can construct the surface plot of the three-phase contact region using the interference patterns.

The sequence of photographs shown in Fig. 4b–e depict the three-phase contact angle dynamics and the various steps of oil droplet removal from the glass surface in the presence of an aqueous micellar solution. (The solution consisted of anionic sodium lauryl sulphate surfactant of an effective micelle diameter of 8 nm, and at a concentration equal to ten times the critical micelle concentration, that is, an effective volume fraction of 0.4, or a concentration of 0.1 mol l⁻¹). We see that the micellar solution penetrates between the oil and the glass surface (Supplementary Information movie 3). The speckled band between the dark and light areas in Fig. 4b and c shows the formation of small aqueous lenses (seen as the white spots

encircled by dark fringes) between the oil and solid surface. In effect, two contact regions are established: the first is between the oil droplet, the solid surface and the aqueous surfactant solution (outer region), and the second is between the oil droplet, the solid surface and the aqueous film with lenses (inner region). A pre-wetted aqueous film with lenses is formed between the two regions. The thickness of the speckled band increases with the lapse of time, because the inner contact region recedes more rapidly than the outer contact line (Fig. 4d). Eventually, the oil droplet is separated completely from the solid surfaces by a thick aqueous film with a dimple (Fig. 4e).

Our explanation for the detachment of an oil drop from the solid surface using a surfactant micellar solution (without an added electrolyte) is that with the lapse of time, micelles diffuse into the wedge film and interact with the film surfaces. The micelle concentration in the film increases. The disjoining pressure increases significantly at a wedge thickness corresponding to one micelle layer (Fig. 3a). As a result of the pressure increase, the oil–solution interface moves forward, and the aqueous micellar solution spreads on the solid surface, detaching the oil drop.

We performed an additional experiment to test the role of micellar interactions in the oil detachment process. In this experiment, we observed the separation of an oil drop from the glass surface in the presence of the same ionic surfactant but with an added electrolyte (0.1 mol l⁻¹ NaCl). However, the drop did not become detached from the solid surface. This was unexpected, because the interfacial tension at the interface between the oil and the aqueous micellar solution decreases at a higher salt concentration and, as a result, the droplet shrinks (Supplementary Information movie 4). This reduction in interfacial tension should have enhanced the separation process. In fact, at high salt concentrations, the effective micellar diameter (and thereby the micelle volume fraction) decreases, owing to the shrinkage of the electrical double layer around each micelle. Therefore, the magnitude of the structural disjoining pressure diminishes, thus reducing the driving force for the detachment of the drop. Further investigation is underway to elucidate the possible effects of salt concentration, micellar concentration and type, and micelle size and polydispersity on the role of

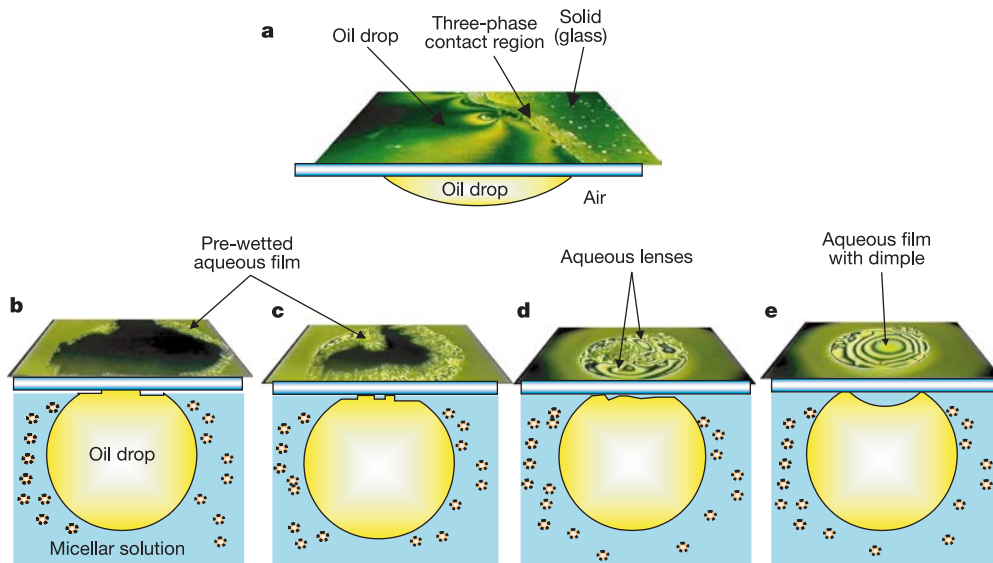


Figure 4 Dynamics of the three-phase contact region. **a**, Photomicrograph showing the oil drop placed on a glass surface and differential interference patterns formed at the three-phase (solid–liquid–air) contact region. **b–e**, Photomicrographs taken at increasing times after addition of the aqueous micellar solution: **b**, 30 s; **c**, 2 min; **d**, 4 min; and **e**, 6 min. **b**, The beginning of the formation of the pre-wetting aqueous

film between the glass surface and the oil droplet. **c**, The spreading of the pre-wetting film. **d**, The pre-wetted film now covers the whole area, and small water lenses are formed. **e**, The separation of oil droplet from the glass surface by a thick aqueous film with a dimple.

the structural disjoining pressure arising from the confined micelles in the oily soil removal process. □

Received 17 December 2002; accepted 21 March 2003; doi:10.1038/nature01591.

1. Brauman, J. I. & Szuromi, P. Thin films. *Science* **273**, 855 (1996).
2. de Gennes, P. G. Wetting: statics and dynamics. *Rev. Mod. Phys.* **57**, 827–863 (1985).
3. Churaev, N. V. On the forces of hydrophobic attraction in wetting films of aqueous solutions. *Colloids Surf. A* **79**, 25–31 (1993).
4. Ruckenstein, E. Effect of short-range interactions on spreading. *J. Colloid Interface Sci.* **179**, 136–142 (1996).
5. Churaev, N. V. et al. The superspreading effect of trisiloxane surfactant solutions. *Langmuir* **17**, 1338–1348 (2001).
6. De Coninck, J., de Ruijter, M. J. & Voué, M. Dynamics of wetting. *Curr. Opin. Colloid Interface Sci.* **6**, 49–53 (2001).
7. Cazabat, A. M., Frayse, N. & Heslot, F. The wetting films. *Colloids Surf.* **52**, 1–8 (1991).
8. Boda, D., Chan, K. Y., Henderson, D., Wasan, D. T. & Nikolov, A. D. Structure and pressure of a hard sphere fluid in a wedge-shaped cell or meniscus. *Langmuir* **15**, 4311–4313 (1999).
9. Tata, B. V. R., Boda, D., Henderson, D., Nikolov, A. D. & Wasan, D. T. Structure of charged colloids under a wedge confinement. *Phys. Rev. E* **62**, 3875–3881 (2000).
10. Trokhymchuk, A., Henderson, D., Nikolov, A. D. & Wasan, D. T. A simple calculation of structural and depletion forces for fluids/suspensions confined in a film. *Langmuir* **17**, 4940–4947 (2001).
11. Wasan, D. T. & Nikolov, A. D. in *Supramolecular Structure in Confined Geometries* (eds Marne, G. & Warr, G.) 40–53 (ACS Symposium Series No. 736, American Chemical Society, Washington DC, 1999).
12. Pieranski, P., Stezleski, L. & Pansu, B. Thin colloidal crystals. *Phys. Rev. Lett.* **50**, 900–903 (1983).
13. Pieranski, P. Colloidal crystals. *Contemp. Phys.* **24**, 25–73 (1983).
14. van Winkel, D. H. & Murray, C. A. Layering in colloidal fluids near a smooth repulsive wall. *J. Chem. Phys.* **89**, 3885–3891 (1988).
15. Murray, C. A. & Gier, D. Colloidal crystals. *Am. Sci.* **83**, 238–245 (1995).
16. Neimark, A. V. & Kornev, K. G. Classification of equilibrium configurations of wetting films on planar substrates. *Langmuir* **16**, 5526–5529 (2000).
17. Kao, R. L., Wasan, D. T., Nikolov, A. D. & Edwards, D. A. Mechanisms of oil removal from a solid surface in the presence of anionic micellar solutions. *Colloids Surf.* **34**, 389–398 (1989).
18. Dimitrov, A. S., Kralchevsky, P. A., Nikolov, A. D. & Wasan, D. T. Contact angles of thin liquid films—interferometric determination. *Colloids Surf.* **47**, 299–321 (1990).
19. Lobo, L. A., Nikolov, A. D., Dimitrov, A. S., Kralchevsky, P. A. & Wasan, D. T. Contact angle of air bubbles attached to an air/water surface in foam applications. *Langmuir* **6**, 995–1001 (1990).
20. Nikolov, A. D., Kralchevsky, P. A., Ivanov, I. B. & Dimitrov, A. S. in *Thin Liquid Film Phenomena* (eds Krantz, W. B., Wasan, D. T. & Jain, R. K.) 82–87 (Symposium Series No. 252, Vol. 82, American Institute of Chemical Engineers, 1986).
21. Nikolov, A. D., Dimitrov, A. S. & Kralchevsky, P. A. Accuracy of the differential–interferometric measurements of curvature—experimental study with liquid drops. *Opt. Acta* **33**, 1359–1386 (1986).
22. Nikolov, A. D. & Wasan, D. T. in *Handbook of Surface Imaging and Visualization* (ed. Hubbard, A.) 209–214 (CRC, Boca Raton, Florida, 1995).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements This work was supported by the National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.T.W. (wasan@iit.edu).

Redistribution of energy available for ocean mixing by long-range propagation of internal waves

Matthew H. Alford

Applied Physics Laboratory and School of Oceanography, University of Washington, 1013 NE 40th Street, Seattle, Washington 98105, USA

Ocean mixing, which affects pollutant dispersal, marine productivity and global climate¹, largely results from the breaking of internal gravity waves—disturbances propagating along the ocean's internal stratification. A global map of internal-wave dissipation would be useful in improving climate models, but would require knowledge of the sources of internal gravity waves and their propagation. Towards this goal, I present here

computations of horizontal internal-wave propagation from 60 historical moorings and relate them to the source terms of internal waves as computed previously^{2,3}. Analysis of the two most energetic frequency ranges—near-inertial frequencies and semidiurnal tidal frequencies—reveals that the fluxes in both frequency bands are of the order of 1 kW m^{-1} (that is, 15–50% of the energy input) and are directed away from their respective source regions. However, the energy flux due to near-inertial waves is stronger in winter, whereas the tidal fluxes are uniform throughout the year. Both varieties of internal waves can thus significantly affect the space-time distribution of energy available for global mixing.

The Earth's meridional overturning circulation (MOC), wherein millions of cubic metres of water sink each second at high latitude and upwell at low, is a vital element of ocean circulation and a key indicator of climate change in models⁴. Theory, profile fits and inverse models of ocean circulation^{5,6} suggest that the MOC requires of the order of 2 TW ($= 2 \times 10^{12} \text{ W}$) of power (though some argue for less⁷) to turbulently warm the abyssal waters so they can upwell. Breaking of internal gravity waves is thought to fuel this turbulence. Sufficient power appears to be available from the combined effects of the wind and the tides^{2,3,5,8,9}, but neither the partitioning of mixing between these sources, nor the energy pathways and resultant mixing distributions, are yet known. Direct measurements^{1,10,11} suggest that abyssal mixing is weak over much of the oceans, but spatially variable. Resolution of these issues is a major goal of physical oceanographers because climate models cannot resolve these processes, but are sensitive to both the magnitude and distribution of the resultant mixing¹². However, directly measuring mixing at all locations is not practical, so another approach is sought here.

Internal gravity waves occupy a frequency range bounded below by the local inertial frequency, $f \equiv 2 \sin(\text{latitude})$ cycles per day. By far the most energetic are 'near-inertial' waves, with frequencies $\omega \approx f$, and internal tides, at the astronomical frequencies (dominantly the semidiurnal M_2). The former are primarily generated by sudden wind events^{9,13} (and, to an unknown degree, by geostrophic adjustment at eddies and fronts); the latter by depth-independent tidal flow over sloping topography.

The energy inputs, S_f and S_{M_2} , into each type of motion vary substantially over the globe. Near-inertial energy input peaks during local winter near storm tracks (western mid-latitudes of each ocean), with a global total of 0.5 TW (ref. 3). This energy is initially input at $\omega \approx f$ into the ocean's surface layer, but models¹⁴ and observations^{15,16} show that deeper disturbances are quickly excited that propagate downward and equatorward towards lower f . The energy available for the M_2 internal tide is maximum near sloping topography, and has a total of 0.7 TW (ref. 2). Together with 1 TW of lower-frequency wind input⁸, there appears to be enough power to provide the 2 TW needed for abyssal stratification maintenance.

If waves did not propagate far, all mixing would occur near the sources. However, internal tides can be observed from space^{17,18} propagating thousands of kilometres from sources. Theory^{19,20} and circumstantial evidence¹⁶ suggest that near-inertial waves can also propagate a long way. Here I address the question 'how much does long-range propagation cause internal-wave dissipation distributions to differ from the sources?' In a steady ocean with perfect knowledge of the sources, $S(\mathbf{x})$, and of the subsequent horizontal energy flux by propagation, $F(\mathbf{x})$, internal-wave dissipation $D(\mathbf{x})$ is given by:

$$D(\mathbf{x}) = S(\mathbf{x}) - \nabla \cdot \mathbf{F}(\mathbf{x}) \quad (1)$$

where $\mathbf{x}$ is position. $\mathbf{F}$ is computed here in each frequency band at 60 discrete locations, using historical moored records and climatological profiles²¹ (Fig. 1, Methods), and compared to previously computed source distributions for each.

Typical time series of depth-integrated near-inertial flux (Fig. 2a, b) show both positive and negative contributions, occurring in pulses of several days' duration and magnitude of the order of $10\text{--}30\text{ kW m}^{-1}$. Magnitudes are several times greater during the two (southern) winters spanned (shaded). The time integral of each (Fig. 2c, d) indicates predominant wintertime propagation toward the northeast, and nearly no net flux during summertime (visible as zero slope), consistent with storm generation at $\omega \approx f$ and subsequent equatorward propagation.

As is the case for energy input by storms^{13,9}, near-inertial energy passage occurs in finite bursts, with a few events accounting for much of the total (for example, June 1991, April 1992). This is borne out in probability density functions (not shown) of all near-inertial flux estimates for all moorings, which exhibit long tails indicative of an intermittent process. These findings support the notion of internal-wave "groups"²². Efforts to identify particular storms associated with these events are under way. By contrast, the tidal-band fluxes (not shown) show more constant energy passage and shorter-tailed distributions, consistent with their more constant forcing.

Near-inertial fluxes in both hemispheres (Fig. 3a) are directed overwhelmingly equatorward, as demonstrated in the histogram of poleward flux (inset), and generally away from regions of large energy input from the wind to near-inertial motions (colour)³. In the North Pacific and North Atlantic, where coverage is denser, fluxes generally decrease in magnitude toward the east, as do the wind-energy inputs. The zonal component of the fluxes is predominantly eastward, as expected given the predominance of eastward-travelling storms¹⁵.

The tidal fluxes (Fig. 3b) are generally directed away from internal-tide sources² (red). Examples in the figure include arrows

emanating from the western Pacific trench, and north across the Equator from the Tuamotu archipelago¹⁷ at 20°S , 220°E and/or Micronesia/Melanesia, further west. Strong fluxes emanate from some continental shelves; for example, northern Europe (also observed in ref. 23) and Cuba. Elsewhere (Brazil, west coast of USA), propagation is on-shelf.

Northeastward energy radiation is observed along 150°W at 25°N and 40°N , in agreement with altimetric estimates^{17,18}, which only detect mode 1, astronomically phase-locked signals. Strong northward fluxes are seen here well past where phase-locked fluxes have weakened, indicating that phase-locked signals may be a small fraction of the total. The weak flux observed further west at the BEMPEX site (41°N , 163°W) here and in ref. 24 may result from superposition of northward fluxes from Hawaii and southward fluxes from the Aleutians, also suggested from altimetric estimates¹⁸.

Coverage is fair in the North Atlantic and Pacific, but very poor at low latitudes and in the Southern Hemisphere, particularly in the Pacific. Inclusion of all suitable data from the archives of Woods Hole and the British Oceanographic Data Centre would improve coverage in the North Atlantic, but would only provide a few more points in the Southern Ocean, and none in the South Pacific. It seems that resolution of fluxes in these oceans must await new deployments (preferably "profiling moorings"²⁵ capable of resolving all modes). Several well-placed moorings would considerably improve the picture.

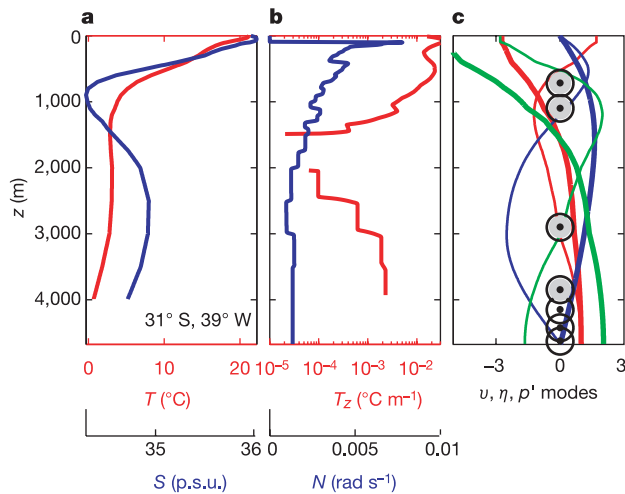


Figure 1 Hydrographic profiles, mode structure and instrument depths for a typical mooring. **a**, Profiles²¹ of climatological temperature (red) and salinity (blue) at the mooring location (31°S , 39°W). **b**, Temperature gradient T_z (red) and buoyancy frequency $N(z)$ (blue) computed from **a**. **c**, Thick and thin lines show respectively first and second vertical normal modes of velocity u (red), vertical displacement η (blue), and pressure anomaly p' (green). Circles indicate instrument depths. In this and all calculations, temperature measurements where $T_z < 3 \times 10^{-5} \text{ }^\circ\text{C m}^{-1}$, or where T increases towards the sea floor, are removed from the calculation (bottom three records here). Briefly, flux is computed as follows (see Methods for details): η is inferred from band-pass-filtered temperature records and climatological profiles (**a**, **b**). A least-squares solution is obtained for the amplitude of the two longest-wavelength normal modes (**c**) in terms of the discrete-depth moored measurements. Depth-integration yields p' in terms of η . The correlation between velocity and p' equals the energy flux.

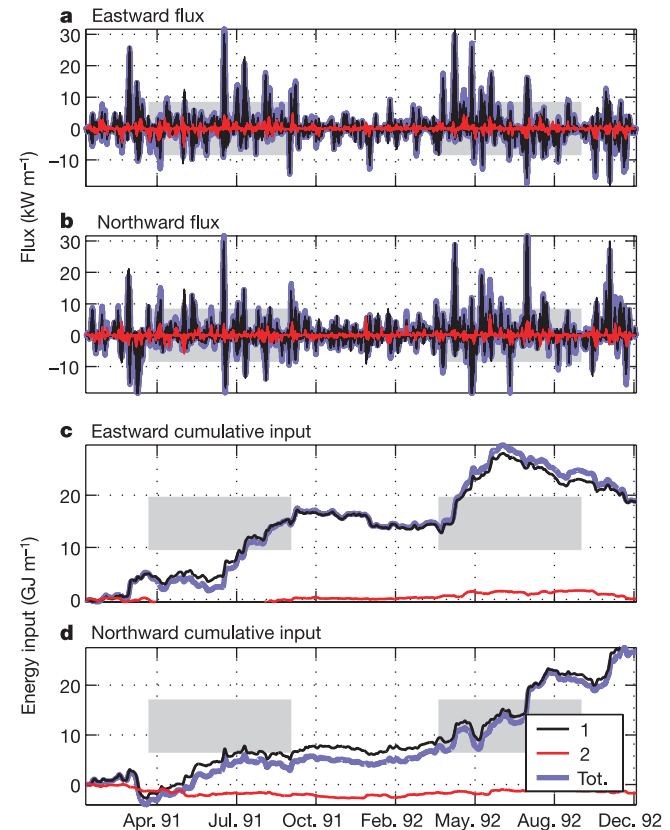


Figure 2 Typical near-inertial time series. **a**, **b**, Depth-integrated eastward (**a**) and northward (**b**) energy flux in the near-inertial band for mode 1 (black), mode 2 (red), and the total (blue), from the mooring in Fig. 1. Southern winter (defined as 1 April–30 September) is shaded. Mean flux in **a** and **b** is respectively 0.32 and 0.45 kW m^{-1} . **c**, **d**, The time-integral of panels **a** and **b**, indicating cumulative energy passage. For this mooring, the flux is wintertime-enhanced, equatorward, and dominated by mode 1.

In spite of the poor spatial coverage, a comparison of near-inertial fluxes to wind-energy inputs is computed for a control volume enclosing the Pacific storm track (Fig. 3a, box). An annual-mean energy input of 65 ± 10 GW (13% of the global total) is computed by integrating the wind-energy flux³ over the box. The five moorings spanning the box's southern edge display a mean southward flux of $\sim 1.5 \text{ kW m}^{-1}$, for a 12-GW energy loss rate due to near-inertial propagation across the box's southern side. Based on these (admittedly few) moorings, at least 15–20% of the energy input propagates significantly from its source region. (Because the horizontal flux estimate may be low (see Methods), and the wind input is an upper bound on energy available for propagating waves, this percentage is a lower bound.) The remainder is dissipated locally either during generation or by breaking before the waves have left the box.

This 12 GW is available to fuel dissipation in the lower latitudes, where the wind forcing is much weaker. The associated breaking, if uniform over the area bounded by $0\text{--}30^\circ \text{N}$, $150\text{--}230^\circ \text{E}$, would yield a depth-mean dissipation rate $\sim 10^{-10} \text{ W kg}^{-1}$, typical of 'background' thermocline mixing rates¹. The horizontal and vertical distribution of the mixing (not resolved) would depend on the waves' interactions and propagation routes.

Similar calculations can be done for tidal propagation across control surfaces (Fig. 3b) parallel to the western Pacific trench and the Hawaiian ridge source regions. Strong tidal flow across abrupt bottom features results in 45 ± 10 GW and 20 ± 5 GW (6.5% and 3% of the global total) of M_2 internal-tidal production at these respective locations². In the western Pacific, the mean flux is

$\sim 2 \text{ kW m}^{-1}$ away from the region, indicating 10 GW ($\sim 25\%$) transmission eastward across the line. For Hawaii, the mean flux is $\sim 1 \text{ kW m}^{-1}$ to the northeast for 3 GW ($\sim 15\%$) transmission across the line. For both volumes, the transmission out the other side is unmeasured; if symmetric, these percentages should be doubled, for 50% and 30%, respectively. Though uncertain, these numbers support other theoretical²⁶, modelling²⁷ and observational (preliminary results from the Hawaii Ocean Mixing Experiment; E. Kunze, personal communication) suggestions that much of internal-tidal energy leaves the source region.

The global internal-wave field has long been viewed as near-inertial and tidal peaks on top of a broadband, isotropic spectral 'continuum'²⁸, whose energy is remarkably constant in both space and time. The near-inertial peak has been argued²⁹ to result from the decrease of f towards the Equator: amplification results when waves propagating from lower latitudes encounter a 'turning latitude' when their frequency equals the local f . The observed equatorward propagation results because waves generated at f must travel equatorward towards lower f , suggesting an alternative explanation of the inertial peak. The present results support the decade-old suggestion¹⁹ that long-distance, anisotropic propagation of near-inertial and tidal internal waves smooths the spatially and temporally variable sources into the constant and isotropic continuum, redistributing energy towards breaking scales via nonlinear interactions. Surface–bottom interactions²⁰ and refraction by eddies and large-scale currents are surely important, but apparently do not completely disrupt propagation.

The present work indicates that low-mode internal-wave energy

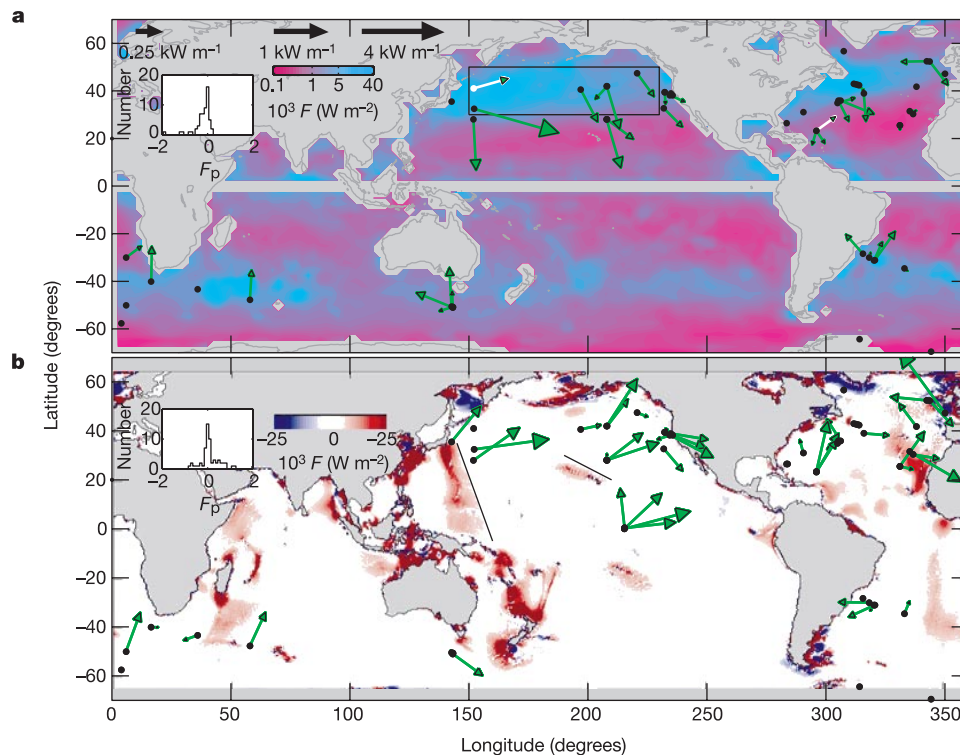


Figure 3 Source terms and energy-flux vectors. **a**, Depth-integrated, annual-mean near-inertial energy-flux vectors for modes 1 and 2 from 60 historical mooring records. The arrow lengths are logarithmically related to flux, with reference lengths indicated at upper left. Moorings with $|F| < 0.1 \text{ kW m}^{-1}$ are shown as black dots without arrows. The few instances of poleward propagation are plotted in white. Colour map indicates annual-mean energy input from the wind to near-inertial mixed-layer motions³. The colour scale is logarithmic, and is indicated at upper left. **b**, As in **a** but for the semidiurnal band. Colour

map denotes internal-tide conversion using the TPX0.5 model² (G. D. Egbert, personal communication). Insets, histogram of the poleward flux component for all moorings. Flux uncertainties (see Methods) are $\pm \sim 50\%$ (magnitude) and $\pm 30^\circ$ (direction). For latitudes $|\lambda| > 50^\circ$ (for example, northern European shelf), the frequency bands in **a** and **b** overlap, preventing their separation. The box and lines in the Pacific are discussed in the text.

fluxes can be computed from suitable mooring records, and that the resulting global flux fields show coherent patterns that are generally directed away from source regions. The fluxes in both the near-inertial and semidiurnal bands are large enough to transport, across ocean basins, globally significant amounts of energy available for mixing. On the basis of the global totals of the source terms^{2,3,9}, the winds and the tides seem able to supply most of the power needed to maintain the deep stratification. Global mapping of their fluxes is an important step in determining how and where internal-wave mixing is accomplished, but the work has just begun. □

Methods

Data selection

I examined data from 1,064 moorings that were deployed from 1973 to 2000 and archived by the Oregon State University Buoy Group (I also thank D. Luther for the BEMPEX mooring data). Only 60 satisfied the following criteria: ≥ 4 instruments of which the first was shallower than 800 m and the deepest $< 1,000$ m from the bottom; sampling interval ≤ 3 h; water depth $H \geq 3,200$ m; record length ≥ 180 d, spanning at least one winter; and no excessive mooring motion (evident as coherent measured temperature and pressure).

Band-pass filtering

Flux was computed separately for the near-inertial and semidiurnal bands via band-pass filtering. A fourth-order Butterworth filter was designed with zero-phase response, and quarter-power points at $\{c^{-1}\omega_c, c\omega_c\}$ around a centre frequency $\omega_c = \{f, M_2\}$. The bandwidth parameter, $c = 1.25$, was chosen to be narrow enough to maximally isolate the two bands, while being wide enough to avoid filter 'ringing'. For this value, motions in the near-inertial and tidal bands are separable for latitudes $|\lambda| < 50^\circ$. Poleward of this, the frequency ranges overlap (that is, $cf > c^{-1}M_2$), and their motions cannot be distinguished.

Computing vertical displacement from temperature

Climatological hydrographic data²¹ were used to compute vertical temperature gradient $T_z(z)$ and buoyancy frequency $N(z)$ at each mooring location (example in Fig. 1a, b). $N^2(z) \equiv g\partial\rho/\partial z$, where g is gravitational acceleration and ρ is potential density, is a measure of the rate with which density increases with depth. Vertical displacement η is computed via the relation

$$\eta(z_j, t) = T(z_j, t)T_z^{-1}(z_j) \quad (2)$$

where $T(z_j, t)$ is the band-passed temperature at depth z_j and time t . This assumes that temperature signals are due to vertical advection of a constant gradient past the mooring. Thus, a cold temperature excursion implies that water has risen from below the instrument, implying $\eta > 0$.

Advection of horizontal gradients by the inertial and tidal currents also contribute to η , but do not influence the mean fluxes: an inertial current of $\pm 0.1 \text{ m s}^{-1}$ at latitude 30° has horizontal displacements of ± 2 km. Horizontal gradients on this scale change sign frequently; therefore, the resultant (spurious) fluxes average out. (Large-scale meridional temperature gradients, which could introduce biases, yield negligible $\eta_{\text{false}} < 1 \text{ m}$.)

Modes

Over a flat bottom, the exact vertical structure of internal waves can be represented by a superposition of discrete vertical modes (Fig. 1c) that depend only on $N(z)$. For low-frequency waves ($\omega \ll N$), these are the solutions to

$$\frac{\partial^2}{\partial z^2}\eta_i(z) + c_i^{-2}N^2(z)\eta_i(z) = 0 \quad (3)$$

subject to the boundary conditions $\eta(0) = \eta(H) = 0$, where c_i are the eigenspeeds. The modal expressions for velocity (u), displacement (η) and pressure anomaly (p') are all related to each other via the internal-wave polarization relations.

Full-depth profiles of modal velocity, $u_i(z, t)$, and displacement, $\eta_i(z, t)$, are obtained, for the first two modes, from the filtered time series $u(z_j, t)$ and $\eta(z_j, t)$ at discrete depths z_j by performing a least-squares inverse at each time t . The resulting time series of modal amplitude typically rotate inertially (clockwise/anticlockwise in the northern/southern hemisphere) with small residual errors.

Flux

The baroclinic pressure anomaly in each mode, p'_i , is obtained by depth-integrating $N^2\eta_i(z)$ and subtracting the mean³⁰. The baroclinic energy flux in each mode is then given by $F_i = \langle u_i p'_i \rangle$, where $\langle \rangle$ indicates average over a wave period. Only the depth-integrated flux, $\int_0^H F_i dz$, which has units of kW m^{-1} , is discussed.

Errors

Energy flux is dominated by the lowest modes³⁰, as in Fig. 2. Higher-mode ($i > 2$) motions are aliased onto the resolved modes, introducing errors. Simulations using various reasonable stratifications, modal content, and instrument locations indicate that for typical four-instrument moorings, the two-mode estimates fall between 65% and 100% of the fully-resolved flux. The flux estimates are therefore (1) biased somewhat low if higher-mode input is present, and (2) uncertain by $\sim 50\%$. (High-mode wave inputs with antinodes near the instrument depths can cause the two-mode solutions to be

overestimates. Highly focused internal-wave rays²³ would therefore be very poorly measured; however, it is unlikely that these are common enough in the deep sea to significantly affect the results). A comprehensive treatment of the method's errors and subtleties will be reported elsewhere (J. Nash, M.H.A. and E. Kunze, manuscript in preparation).

Because tidal forcing varies little (except though varying background stratification and currents), the spread in tidal-flux estimates at collocated moorings are another estimate of the method's intrinsic uncertainty. Conservative values of $\pm \sim 50\%$ and $\pm 30^\circ$ are taken for the magnitude and directional uncertainties, consistent with the error estimates from the simulations. (The four moorings on the Equator show similar magnitudes but considerable angular spread. Here, refraction by equatorial currents may be stronger.)

Received 22 October 2002; accepted 4 April 2003; doi:10.1038/nature01628.

- Gregg, M. Diapycnal mixing in the thermocline. *J. Geophys. Res.* **92**, 5249–5286 (1987).
- Egbert, G. D. & Ray, R. D. Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data. *Nature* **405**, 775–778 (2000).
- Alford, M. H. Improved global maps and 54-year history of wind-work on ocean inertial motions. *Geophys. Res. Lett.* **30**(8), doi:10.1029/2002GL016614 (2000).
- Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1**, 841–866 (1988).
- Munk, W. & Wunsch, C. Abyssal recipes II: energetics of tidal and wind mixing. *Deep-Sea Res.* **1**, 1977–2010 (1998).
- Ganachaud, A. & Wunsch, C. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature* **408**, 453–456 (2000).
- Webb, D. J. & Sugimoto, N. Vertical mixing in the ocean. *Nature* **409**, 37–39 (2001).
- Wunsch, C. The work done by the wind on the oceanic general circulation. *J. Phys. Oceanogr.* **28**, 2332–2340 (1998).
- Alford, M. H. Internal swell generation: the spatial distribution of energy flux from the wind to mixed-layers near-inertial motions. *J. Phys. Oceanogr.* **31**, 2359–2368 (2001).
- Ledwell, J. R., Watson, A. J. & Law, C. S. Evidence for slow mixing across the pycnocline from an open-ocean tracer-release experiment. *Nature* **364**, 701–703 (1993).
- Polzin, K. L., Toole, J. M., Ledwell, G. R. & Schmitt, R. W. Spatial variability of turbulent mixing in the abyssal ocean. *Science* **276**, 93–96 (1997).
- Samelson, R. M. Large-scale circulation with locally enhanced vertical mixing. *J. Phys. Oceanogr.* **28**, 712–726 (1998).
- D'Asaro, E. The energy flux from the wind to near-inertial motions in the mixed layer. *J. Phys. Oceanogr.* **15**, 943–959 (1985).
- Gill, A. E. On the behavior of internal waves in the wake of a storm. *J. Phys. Oceanogr.* **14**, 1129–1151 (1984).
- D'Asaro, E. *et al.* Upper-ocean inertial currents forced by a storm, I, data and comparisons with linear theory. *J. Phys. Oceanogr.* **25**, 2909–2936 (1995).
- Alford, M. & Gregg, M. Near-inertial mixing: modulation of shear, strain and microstructure at low latitude. *J. Geophys. Res.* **106**, 16947–16968 (2001).
- Ray, R. D. & Mitchum, G. T. Surface manifestations of internal tides in the deep ocean: observations from altimetry and island gauges. *Prog. Oceanogr.* **40**, 135–162 (1997).
- Ray, R. D. & Cartwright, D. E. Estimates of internal tide energy fluxes from TOPEX/POSEIDON altimetry: central North Pacific. *Geophys. Res. Lett.* **28**, 1259–1262 (2001).
- D'Asaro, E. A. In *Proc. Dynamics of Oceanic Internal Gravity Waves II* (eds Müller, P. & Henderson, D.) 451–466 ('Aha Huli'ko'a Hawaiian Winter Workshop, Hawaii Institute of Geophysics, Honolulu, 1991).
- Garrett, C. What is the "near-inertial" band and why is it different from the rest of the internal wave spectrum? *J. Phys. Oceanogr.* **31**, 962–971 (2001).
- Levitus, S. & Boyer, T. *World Ocean Atlas 1994* (NOAA Atlas NESDIS 4, US Department of Commerce, Washington DC, 1994).
- Thorpe, S. A. On internal wave groups. *J. Phys. Oceanogr.* **29**, 1085–1095 (1999).
- Pingree, R. D. & New, A. L. Abyssal penetration and bottom reflection of internal tidal energy in the Bay of Biscay. *J. Phys. Oceanogr.* **21**, 28–39 (1991).
- Dushaw, B., Cornuelle, B., Worcester, P. F., Howe, B. & Luther, D. Barotropic and baroclinic tides in the Central North Pacific Ocean determined from long-range reciprocal acoustic transmissions. *J. Phys. Oceanogr.* **25**, 631–647 (1995).
- Doherty, K. W., Frye, D. E., Liberatore, S. P. & Toole, J. M. A moored profiling instrument. *J. Atmos. Ocean Technol.* **16**, 1816–1829 (1999).
- St Laurent, L. & Garrett, C. The role of internal tides in mixing the deep ocean. *J. Phys. Oceanogr.* **32**, 2882–2899 (2001).
- Merrifield, M. A. & Holloway, P. E. Model estimates of M_2 internal tide energetics at the Hawaiian Ridge. *J. Geophys. Res.* **107**, doi:10.1029/2001JC000996 (2002).
- Garrett, C. J. R. & Munk, W. H. Space-time scales of internal waves: a progress report. *J. Geophys. Res.* **80**, 291–297 (1975).
- Fu, L. L. Observations and models of inertial waves in the deep ocean. *Rev. Geophys. Space Phys.* **19**, 141–170 (1981).
- Kunze, E., Rosenfield, L., Carter, G. & Gregg, M. C. Internal waves in Monterey submarine canyon. *J. Phys. Oceanogr.* **32**, 1890–1913 (2002).

Acknowledgements I thank E. D'Asaro, E. Kunze and C. Garrett for insights and discussions, and M. Whitmont for quality control and preparation of the mooring data. This work was supported by the Office of Naval Research Young Investigator award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to the author (malford@apl.washington.edu).

Experimental evidence that potassium is a substantial radioactive heat source in planetary cores

V. Rama Murthy*, Wim van Westrenen†‡ & Yingwei Fei†

* Department of Geology and Geophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA and Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA

† Geophysical Laboratory, Carnegie Institution of Washington, Washington DC 20015, USA

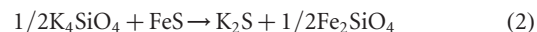
‡ Institute for Mineralogy and Petrography, ETH Zürich, 8092 Zürich, Switzerland

The hypothesis that ^{40}K may be a significant radioactive heat source in the Earth's core was proposed on theoretical grounds^{1,2} over three decades ago, but experiments^{3–8} have provided only ambiguous and contradictory evidence for the solubility of potassium in iron-rich alloys. The existence of such radioactive heat in the core would have important implications for our understanding of the thermal evolution of the Earth and global processes such as the generation of the geomagnetic field, the core–mantle boundary heat flux and the time of formation of the inner core^{9–12}. Here we provide experimental evidence to show that the ambiguous results obtained from earlier experiments are probably due to previously unrecognized experimental and analytical difficulties. The high-pressure, high-temperature data presented here show conclusively that potassium enters iron sulphide melts in a strongly temperature-dependent fashion and that ^{40}K can serve as a substantial heat source in the cores of the Earth and Mars.

We have found that polishing of a sample for electron microprobe analysis using oil lubricants results in substantial K loss from the Fe–S phase within a few hours and a near total loss in a few days (Fig. 1). A second problem is the loss of K in experiments when the charge is contained in a single capsule, as shown by mass-balance calculations after the experiment. The degree of loss depends on temperature (T), run duration and pressure (P). These unrecognized problems in previous studies could have contributed to the ambiguities concerning K entry into iron sulphide melts. The validity of other data on alkali distribution in terrestrial and

planetary materials gathered by such experiments may also need to be examined.

We report here data from experiments specifically designed to avoid these analytical difficulties (see Methods). Our initial effort was to establish that reactions such as those suggested previously², for example:



are possible and that K_2S is soluble in a Fe–S liquid that would have segregated to form the core.

Synthetic mixtures of Fe–metal, FeS, a K–silicate glass and, in one instance, a natural peridotite (KLB-1) were used for experiments conducted at 1–3 GPa at temperatures above the liquidus of both metal and silicate phases (1,200–1,800 °C). The calculated oxygen fugacity, f_{O_2} , in all runs was about 1.5 log units below the iron–wüstite buffer. This is in accord with core-formation models in which metallic liquid equilibrates with molten silicate under reducing conditions in a magma ocean in the early Earth¹³. We determined the partition coefficient for potassium, D_K (concentration of K in sulphide/concentration of K in silicate) as a function of pressure, temperature, and composition.

Data on D_K at different temperatures at constant pressure (2 GPa), oxygen fugacity, and composition are shown in Table 1. We note the strong dependence of D_K on temperature under these conditions. In a plot of $\ln D_K$ versus $1/T$, the data show a precise linear relationship (Fig. 2). These data provide a definitive demonstration that K can be accommodated in Fe–S melts that equilibrated with silicates, eliminating the decades-old experimental ambiguity regarding the entry of K into Fe–S liquids.

Besides temperature, many other variables, such as pressure, composition, f_{O_2} , and melt structure can affect metal–silicate partition coefficients^{14–17}. As already noted, the f_{O_2} in our experiments is in conformity with that expected in core-forming conditions. Previous experiments suggested a significant increase in D_K as the silicate melt becomes more depolymerized (higher number of non-bridging oxygen atoms per tetrahedral cation, NBO/T), although the data showed considerable scatter^{6,7}. Initial experiments in our study, when corrected to a common temperature, instead indicate a nonlinear decrease in D_K with increasing NBO/T in a range of compositions from <1 to a peridotitic composition with NBO/T = 2.7 (Table 2 and Fig. 3). Gessmann and Wood⁷ noted a small negative pressure dependence of D_K , but our experiments, covering a limited pressure range, show no measurable effect (Table 3 and Fig. 4).

Current estimates of the pressure and temperature conditions at which Earth's core materials equilibrated with the mantle fall in the range 25–60 GPa and 2,400–4,200 K (refs 18, 19). Extrapolation from our temperature-dependence data at low NBO/T (Fig. 2) to a plausible range of 3,000–4,000 K yields a D_K value of 1.8–4.6. Extrapolation of this D_K range to account for melt structure, composition and pressure applicable to core formation also needs to be considered. If we consider the dependencies of D_K on NBO/T and pressure from Figs 3 and 4, the applicable D_K for the Earth is in the range 0.2–0.5. These values are sufficiently large to at least qualitatively indicate that a significant amount of K can be sequestered into a S bearing core.

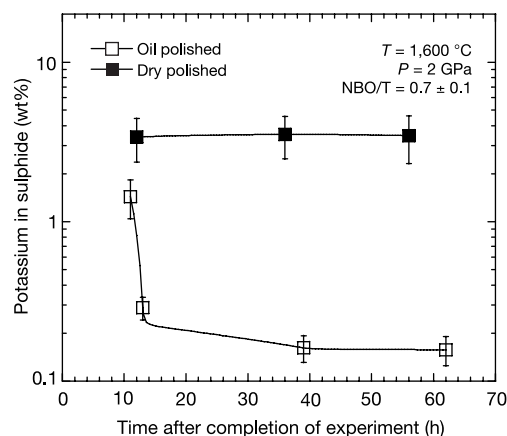


Figure 1 Potassium loss from oil polished sample compared to dry polished sample. Note the substantial K loss in about 30 min and nearly an order-of-magnitude loss in 40 h in a sample that was polished several times in oil. Dry polish yields consistent and higher values with no perceptible loss up to at least 60 h. Error bars are 2σ .

Table 1 Effect of temperature on D_K

Run no.	Run T (°C)	%K in sulphide	%K in silicate	D_K
PC 328-01	1,200	0.67 ± 0.18	16.85 ± 0.22	0.040 ± 0.010
PC 329-01	1,400	1.91 ± 0.43	18.28 ± 0.24	0.10 ± 0.024
PC 331-01	1,500	2.14 ± 1.01	16.12 ± 0.53	0.13 ± 0.062
PC 327-01	1,600	3.66 ± 1.16	17.52 ± 0.79	0.21 ± 0.067

Shown are the concentration of K in quenched liquid Fe–S and silicate phases and D_K as a function of temperature at constant pressure (2.0 GPa) and composition. Errors are 2σ .

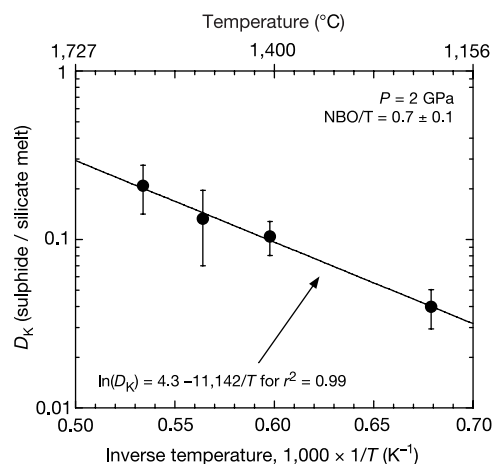


Figure 2 Effect of temperature on D_K . D_K sulphide–silicate distribution coefficients at constant pressure (2 GPa) and composition (NBO/T = 0.7) increase significantly with increasing temperature. Data from Table 1; error bars are 2σ .

Knowledge of the S content of the core is also crucial to establish its K content. The density deficiency of the core suggests about 8–10 wt% S if S is the only light element in the core, but cosmochemical models based on the volatility of S have argued that the S content of the core may only be 2–5 wt% (ref. 19). We note here that in the materials that accreted to form the terrestrial planets, S is stabilized as FeS and therefore, cosmochemical arguments based on volatility of elemental S may not be appropriate. The only direct experimental data bearing on the S content of the Earth's core is the recent study²⁰ that concluded that the S content of metallic liquid in equilibrium with a magma ocean with 200 p.p.m. S is 5.5–12 wt%. Assuming that experiment takes precedence over hypothesis, for a 250 ± 50 p.p.m. S content in the mantle (ref. 21) a core with about 10 wt% of S seems plausible.

Given these considerations, it is possible to evaluate the role of radiogenic heat due to ^{40}K in the Earth's core. The Fe–S phase in our experiments contained about 25 wt% of S. Using the extrapolated D_K range of 0.2–0.5 discussed earlier, the K content of the core is 60–130 p.p.m. with a present-day heat production of 0.4–0.8 TW. Some recent estimates of the core–mantle boundary (CMB) heat flux are in the range of 8–10 TW (refs 11, 22). The radiogenic heat in the core could thus be a significant fraction of the CMB heat flux in this scenario. With this additional heat in the core, the Urey number (total heat production in the Earth divided by heat flow at the surface) will be higher than is presently thought, and might be more

Table 2 Effect of silicate melt composition on D_K

Run no.	Run T (°C)	NBO/T	D_K at run T	D_K at 1,800 °C
PC 331-01	1,500	0.71	0.13 ± 0.062	0.33 ± 0.15
PC 345-01	1,500	0.95	0.048 ± 0.018	0.12 ± 0.045
PC 338-02	1,500	1.24	0.021 ± 0.012	0.052 ± 0.031
PC 335-02	1,300	1.42	0.0059 ± 0.0032	0.033 ± 0.018
PC 333-02	1,800	2.72	0.022 ± 0.013	0.022 ± 0.013

D_K is shown as a function of the number of non-bridging oxygens per tetrahedral cation in the silicate melt (NBO/T) at constant pressure (2 GPa) and temperature (1,800 °C). Experiment PC 333-02 contained, in addition to K–silicate glass, natural KLB-1 peridotite. D_K values at the listed run temperatures were converted to a common temperature of 1,800 °C using the trend shown in Fig. 1. Errors are 2σ .

Table 3 Effect of pressure on D_K

Run no.	P (GPa)	%K in sulphide	%K in silicate	D_K
PC 348-02	1.0	1.21 ± 0.34	20.24 ± 0.18	0.060 ± 0.017
PC 345-01	2.0	1.06 ± 0.40	22.00 ± 0.12	0.048 ± 0.018
PC 346-02	3.0	1.61 ± 0.36	23.54 ± 0.17	0.068 ± 0.015

Concentration of K in quenched liquid Fe–S and silicate phases and D_K as function of pressure at constant temperature (1,500 °C) and composition. Errors are 2σ .

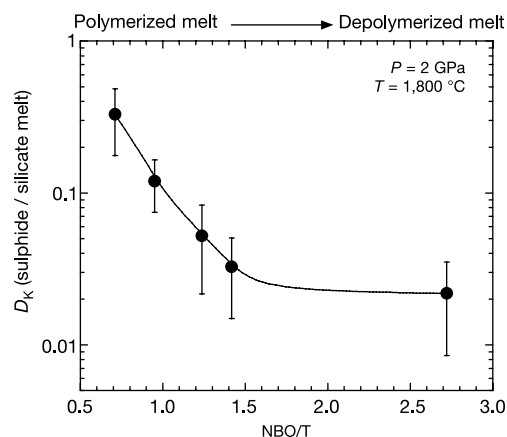


Figure 3 Effect of silicate depolymerization on D_K . At constant temperature and pressure, D_K decreases significantly as the silicate melt becomes increasingly depolymerized (as indicated by an increasing the number of non-bridging oxygens per tetrahedral cation, NBO/T). The NBO/T ratio value relevant to core–mantle equilibrium in the Earth and Mars is around 2.8. Data are from Table 2; error bars are 2σ .

satisfactory for models of whole-mantle convection²³.

Recent studies have indicated that it is difficult to maintain a core dynamo for the past 3.5 Gyr ago (as indicated by the antiquity of the Earth's magnetic field) at the present levels of CMB heat flux, if the heat is due solely to core cooling and solidification of the inner core^{11,22}. The additional ^{40}K radiogenic heat suggested here would alleviate this problem and permit the existence of the magnetic field before 3 Gyr ago.

That significant K can be dissolved in Fe–S liquids even at relatively low-pressure, low-temperature conditions has important implications for Mars with a core richer in S than the Earth's core²⁴. Because of its smaller size, the temperature and pressure of equilibration between silicates and core material in Mars will be lower than in the Earth. A core–mantle equilibration temperature of 2,000–2,500 K yields a D_K of around 0.1. Using a S-rich composition consistent with the moment of inertia and a core of about 15% by mass of the planet^{24,25}, the heat generation due to ^{40}K in the core of Mars would be about 3×10^{10} W at present, and exponentially more so 4 Gyr ago, if its mantle contains about 700 p.p.m. K (ref. 26). Many of the SNC (Shergotty–Nakhla–Chassigny) meteorites presumed to be from Mars contain $\sim 1,000$ p.p.m. K. This level of radiogenic heat will aid an early martian dynamo driven by thermal convection alone to produce the inferred global field in its early history²⁷. The problem is how to stop the dynamo shortly thereafter. But an initially hot core may have cooled quickly to a level at which

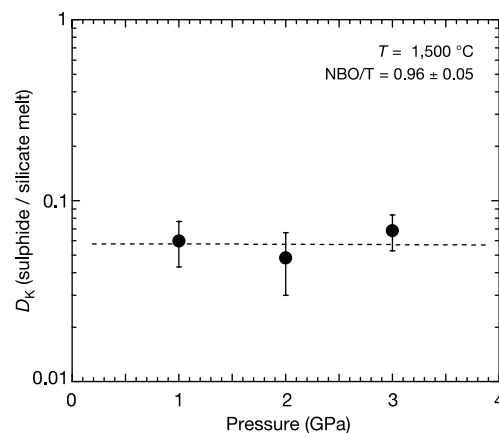


Figure 4 Effect of pressure on D_K . At constant temperature and silicate melt composition, D_K does not vary significantly with pressure. Data are from Table 3; error bars are 2σ .

its heat can be carried by conduction alone and in the absence of convection the dynamo ceases to exist^{12,28,29}. Because of the high S content of the core and the low Fe–FeS eutectic temperature at the pressures applicable to Mars, it is possible that Mars has a totally liquid core at present with no inner core. In any case, radioactive heat generation by ⁴⁰K in the core of Mars should be considered in dynamo models for Mars and its ancient global magnetic field.

In conclusion, it definitely seems possible that radioactive heat due to ⁴⁰K can affect the thermal evolution and global processes in the Earth and Mars. Experimental verification using these techniques would enable us to determine whether U and Th are also potential contributors to radioactive heat in the cores of these planets. □

Methods

Dry polishing of samples

It has long been known that potassium sulphide is highly water-soluble, and previous studies have carefully avoided the use of water during polishing. As an alternative, water-free commercial lapping oils have been routinely employed in recent potassium partitioning experiments. We found that polishing in lapping oil causes significant, time-dependent, and variable loss of potassium from the metal sulphide phase, up to an order of magnitude, depending on how many polishing cycles are used. K is highly labile from the sulphide phase in such samples, and can be completely lost from the sulphide phase within a few hours of exposure to air after oil polish. Here we have developed a dry-polishing technique that avoids all liquids. Hexagonal boron nitride powder, which is nearly twice as lubricious as graphite, is used as a lubricant in the lapping operations. The boron nitride powders we used were Grade HCJ-48 (Union Carbide Corp.) for coarse polishing steps and Grade AC6003 (Advanced Ceramic Corp.) for below 12-μm level polishing. Boron nitride powder spread on carborundum papers allowed good hand polishing of the samples. Excellent polish down to 0.3 μm can be attained by this dry polishing technique and may be particularly useful in studies of distribution of alkali elements in high-pressure, high-temperature experiments.

High-pressure experiments

All experiments reported here used a double capsule technique in which the charge was loaded into an inner graphite capsule that was then sealed in a welded Pt outer capsule. Mass-balance calculations show no loss of K in these experiments, in contrast to those with single capsules. Experiments were conducted in an end-loaded piston cylinder apparatus. Temperatures were monitored using a tungsten-rhenium thermocouple (W5%Re/W26%Re) thermocouple, ignoring the effects of pressure on thermocouple electromotive force. Samples were analysed on the JEOL 8900 electron microprobe at the Geophysical Laboratory. All samples contained blebs of iron sulphide, up to 800 μm in diameter with small amounts (<4% vol.) of pure Fe blebs (<20 μm in diameter), surrounded by glassy silicate. Silicate glasses were analysed at 15 kV and 12 nA using a broad beam of approximately 15 μm in diameter to prevent K volatilization during analysis. Sulphides were analysed at 15 kV and 20 nA with a beam of 5 μm in diameter, using Si as a monitor for silicate contamination. Further analytical details and images of a representative sample are provided in the Supplementary Information.

Received 28 October 2002; accepted 7 March 2003; doi:10.1038/nature01560.

- Lewis, J. S. Consequences of the presence of sulfur in the core of the Earth. *Earth Planet. Sci. Lett.* **11**, 130–134 (1971).
- Hall, H. T. & Murthy, V. R. The early chemical history of the Earth: some critical elemental fractionations. *Earth Planet. Sci. Lett.* **11**, 239–244 (1971).
- Oversby, V. M. & Ringwood, A. E. Potassium distribution between metal and silicate and its bearing on the occurrence of potassium in the Earth's core. *Earth Planet. Sci. Lett.* **14**, 345–347 (1972).
- Goettel, K. A. Partitioning of potassium between silicates and sulphide melts: experiments relevant to the Earth's core. *Phys. Earth Planet. Inter.* **6**, 161–166 (1972).
- Murrell, M. T. & Burnett, D. S. Partitioning of K, U, and Th between sulfide and silicate liquids: Implications for radioactive heating of planetary cores. *J. Geophys. Res.* **91**, 8126–8136 (1986).
- Chabot, N. L. & Drake, M. J. Potassium solubility in metal: the effects of composition at 15 kbar and 1900 °C on partitioning between iron alloys and silicate melts. *Earth Planet. Sci. Lett.* **172**, 323–335 (1999).
- Gessmann, C. K. & Wood, B. J. Potassium in the Earth's core? *Earth Planet. Sci. Lett.* **200**, 63–78 (2002).
- Ohtani, E., Yurimoto, H. & Seto, S. Element partitioning between metallic liquid, silicate liquid, and lower-mantle minerals: implications for core formation of the Earth. *Phys. Earth Planet. Inter.* **100**, 97–114 (1997).
- Gubbins, D., Masters, T. G. & Jacobs, J. A. Thermal evolution of the Earth's core. *Geophys. J. R. Astron. Soc.* **59**, 57–99 (1979).
- Buffett, B. A., Huppert, H. E., Lister, J. R. & Woods, A. W. On the thermal evolution of the Earth's core. *J. Geophys. Res.* **101**, 7989–8006 (1996).
- Labrosse, S., Poirier, J.-P. & Le Mouél, J.-L. The age of the inner core. *Earth Planet. Sci. Lett.* **190**, 111–123 (2001).
- Stevenson, D. J., Spohn, T. & Schubert, G. Magnetism and thermal evolution of the terrestrial planets. *Icarus* **54**, 466–489 (1983).
- Walter, M. J. & Thibault, Y. Partitioning of tungsten and molybdenum between metallic liquid and silicate melt. *Science* **270**, 1186–1189 (1995).
- Jana, D. & Walker, D. The influence of silicate melt composition on distribution of siderophile elements among metal and silicate liquids. *Earth Planet. Sci. Lett.* **150**, 463–472 (1997).
- Righter, K., Drake, M. J. & Yaxley, G. Prediction of siderophile element metal/silicate partition

coefficients to 20 GPa and 2800 °C: the effects of pressure, temperature, oxygen fugacity, and silicate and metallic melt composition. *Phys. Earth Planet. Inter.* **100**, 115–134 (1997).

- Righter, K. & Drake, M. J. Metal/silicate equilibrium in a homogeneously accreting Earth: New results for Re. *Earth Planet. Sci. Lett.* **146**, 541–554 (1997).
- Gessmann, C. K. & Rubie, D. C. The origin of the depletions of V, Cr and Mn in the mantles of the Earth and Moon. *Earth Planet. Sci. Lett.* **184**, 95–107 (2000).
- Li, J. & Agee, C. B. The effect of pressure, temperature, oxygen fugacity and composition on partitioning of nickel and cobalt between liquid Fe–Ni–S alloy and liquid silicate: Implications for the Earth's core formation. *Geochim. Cosmochim. Acta* **65**, 1821–1832 (2001).
- Hillgren, V. J., Gessmann, C. K. & Li, J. in *Origin of the Earth and Moon* (eds Canup, R. M. & Righter, K.) 245–263 (Univ. Arizona Press, Tucson, 2000).
- Holzheid, A. & Grove, T. L. Sulfur saturation limits in silicate melts and their implications for core formation scenarios for terrestrial planets. *Am. Mineral.* **87**, 227–237 (2002).
- McDonough, W. F. & Sun, S.-S. The composition of the Earth. *Chem. Geol.* **120**, 223–253 (1995).
- Buffett, B. A. Estimates of heat flow in the deep mantle based on the power requirements for the geodynamo. *Geophys. Res. Lett.* **29**, doi:10.1029/2001GL014649 (2002).
- McKenzie, D. & Richter, F. M. Parameterized thermal convection in a layered region and the thermal history of the Earth. *J. Geophys. Res.* **86**, 11667–11680 (1981).
- Bertka, C. M. & Fei, Y. Implications of Mars Pathfinder data for the accretion history of the terrestrial planets. *Science* **281**, 1838–1840 (1998).
- Fei, Y., Li, J., Bertka, C. M. & Prewitt, C. T. Structure type and bulk modulus of Fe₂S, a new iron-sulfur compound. *Am. Mineral.* **85**, 1830–1833 (2000).
- Lodders, K. & Fegley, B. An oxygen isotope model for the composition of Mars. *Icarus* **126**, 373–394 (1997).
- Acuña, M. H. *et al.* Global distribution of crustal magnetization discovered by the Mars Global Surveyor MAG/ER experiment. *Science* **284**, 790–793 (1999).
- Stevenson, D. J. Mars core and magnetism. *Nature* **412**, 214–219 (2001).
- Nimmo, F. & Stevenson, D. J. The influence of early plate tectonics on the thermal evolution and magnetic field of Mars. *J. Geophys. Res.* **105**, 11969–11979 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Drake and D. Draper for comments and suggestions, N. Irvine for providing the K–silicate glass, B. Mysen for advice on silicate melt properties, C. Hadidiacos for assistance in electron microprobe analyses and C. Alexander, R. Carlson, E. Hauri, J. Li, F. Nimmo, D. Stevenson and J. Van Orman for discussions. This work was supported by a Grant-in-aid from the University of Minnesota (V.R.M.), the NASA (Y.F.), a Carnegie Institution Postdoctoral Fellowship and the Swiss National Science Foundation (W.v.W.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to V.R.M. (vrmurthy@umn.edu).

Universal scaling relations in food webs

Diego Garlaschelli*†, Guido Caldarelli* & Luciano Pietronero*‡

* INFN UdR Roma 1 and Dipartimento di Fisica Università di Roma 'la Sapienza', P. le A. Moro 5, 00185 Rome, Italy

† INFN UdR Siena and Dipartimento di Fisica Università di Siena, Via Roma 56, 53100 Siena, Italy

‡ CNR, Istituto di Acustica 'O.M. Corbino', v. Fosso del Cavaliere 100, 00133 Roma, Italy

The structure of ecological communities is usually represented by food webs^{1–3}. In these webs, we describe species by means of vertices connected by links representing the predations. We can therefore study different webs by considering the shape (topology) of these networks^{4,5}. Comparing food webs by searching for regularities is of fundamental importance, because universal patterns would reveal common principles underlying the organization of different ecosystems. However, features observed in small food webs^{1–3,6} are different from those found in large ones^{7–15}. Furthermore, food webs (except in isolated cases^{16,17}) do not share^{18,19} general features with other types of network (including the Internet, the World Wide Web and biological webs). These features are a small-world character^{4,5} and a scale-free (power-law) distribution of the degree^{4,5} (the number of

links per vertex). Here we propose to describe food webs as transportation networks²⁰ by extending to them the concept of allometric scaling^{20–22} (how branching properties change with network size). We then decompose food webs in spanning trees and loop-forming links. We show that, whereas the number of loops varies significantly across real webs, spanning trees are characterized by universal scaling relations.

To analyse the statistical features of food webs, we propose to focus on their transportation properties²⁰. This is possible because the directionality of the links (pointing from prey to predator¹) defines a ‘flow’ of resources^{1,23} (energy, nutrients and prey) between the vertices of the web. Here vertices are trophic species^{12,24}, defined as sets of biological species sharing the same predators and the same prey. Because every species feeds directly or indirectly on environmental resources to survive^{1–3}, food webs are connected (that is, every species can be reached by starting from an additional ‘source’ vertex representing the environment²³). This allows us to define a spanning tree of a generic food web as a loopless subset of the links of the web such that, starting from the environment, every species can be reached by the flow (see Fig. 1). Hereafter, the term ‘loop’ indicates any closed path of links, ignoring the direction. A spanning tree of a web with S species (environment excluded) and L links has only S links. The remaining $L - S$ links are a measure of the number of loops in the whole web. For a given web, several spanning trees might exist. In Methods we discuss some possible criteria (chain length minimization and chain strength maximization) that allow the isolation of the minimal spanning tree, which is somehow related to the main transfer of resources in the webs. For reasons of brevity, we denote as ‘strong’ links the S links in the spanning tree and as ‘weak’ links the remaining $L - S$ ones.

Having reduced the complexity of the whole structure to a simpler tree, we can now study how the resources are delivered in the ecosystem. We then perform a typical analysis²⁰ introduced for other transportation networks such as river basins^{20,25} and vascular systems^{20–22}. This analysis consists of measuring the degree of optimization of the transfer rate. In more detail, for each species i in the spanning tree (the environment is labelled by $i = 0$), we compute the number A_i of species feeding directly or indirectly on i (plus species i itself). We also compute the sum $C_i = \sum_k A_k$, where k runs over the set of predators of i plus i itself (see Fig. 1b). Note that $A_0 = S + 1$, where S is the number of species (environment excluded). In both river²⁰ and vascular^{21,22} networks, the shape of C_i as a function of A_i is universal (for every river and any living organism, respectively) and optimized. In Fig. 2a–g, we show the result (see Methods) of plotting C_i as a function of A_i (for each species $i = 0, S$) for the seven largest food webs in the literature, including (in order of increasing number of trophic species S) St Martin Island⁷ ($S = 42$), St Marks seagrass⁸ ($S = 48$), grassland⁹ ($S = 63$), Silwood Park¹⁰ ($S = 81$), Ythan Estuary without para-

sites¹¹ ($S = 81$), Little Rock Lake¹² ($S = 93$) and Ythan Estuary with parasites¹³ ($S = 123$). Remarkably, in all food webs, we find a power-law scaling of the form $C \propto A^\eta$ with exponents in the range 1.13–1.16 (see Fig. 2a–g). These power-law relations are usually called allometric^{20–22}. They describe how the topological properties scale with network size, and are related to the self-similarity^{21,22,25} of the tree-like structure. In other words, they signal that the structure of the whole tree is statistically equivalent to that of any of its branches. Here, the observation of self-similarity is reinforced by the fact that source webs (for example, Silwood Park, reporting all the interactions based on the scotch broom *Cytisus scoparius*¹⁰ and hence being a ‘branch’ of a larger, undocumented web) display the same scaling as the other webs. In a river network²⁵, where A_i is the area of the drainage basin of point i , the exponent is $\eta = 3/2$. This relation holds for any river (the exponent is universal) and hence suggests a common hydrogeological evolution²⁵. For the vascular system^{21,22} of an organism, there is no direct measure of the single values of A_i and C_i . Nevertheless, it is possible to relate the global quantities A_0 and C_0 to the metabolic rate B and to the body mass M , respectively, of the organism²⁰. From Kleiber’s empirical law²⁶, $B(M) \propto M^{3/4}$ (valid for several orders of magnitude of M), we have the relation $C_0 \propto (A_0)^\eta$ with $\eta = 4/3$. In this case, too, the universality of the exponent suggests that vascular systems of all organisms have been shaped by a common evolutionary process²².

Similarly, the observation of universality in food webs would highlight common organizing principles across different ecosystems (despite the diversity^{7–15} of the environments, ranging from terrestrial to aquatic, from desert to freshwater, and so on). We stress that, even in an ideal system, the correct scaling relations are expected to hold in the large-scale limit. Here, the largest webs have an exponent η of 1.13, whereas the smallest ones have higher exponents marginally consistent with 1.13. This indicates that $\eta = 1.13$ might represent the correct (universal) behaviour for large webs. To test this universality hypothesis, we select only the large-scale behaviour for each web. As with metabolic rates, we plot C_0 against A_0 (see Fig. 2h) for the seven webs together (see Methods) plus the webs of Coachella Valley¹⁴ ($S = 29$), Skipwith Pond¹⁵ ($S = 25$) and an average over the 181 webs of the EcoWeb database⁶ (mean $S = 16$). This rather comprehensive analysis includes almost all published food webs; the numbers of species range from $S = 16$ to $S = 124$. Remarkably (see Fig. 2h), we find again a scaling relation fitted by the exponent $\eta = 1.13 \pm 0.03$, confirming the expectation of our universality hypothesis and indicating that the scaling might indeed be invariant across large food webs.

A quantitative comparison with other systems is again particularly revealing. In general, the value of the scaling exponent η measures the efficiency of resource transfer²⁰. In a transportation system, A_i is proportional to the quantity of resources (water or blood) exchanged at point i , and C_i measures the ‘cost’ of this transfer²⁰. Note that here we are referring to a ‘global’ notion of efficiency, measuring whether network topology optimizes the supply of resources to (or from) a given point (the outlet, the heart or the environment) from (or to) any other. This global efficiency should not be confused with the ‘local’ efficiency of a single-link transfer, which is always small^{1–3}, as discussed in Methods. Two independent proofs exist^{20,22} that, for a tree-like network embedded in a space of euclidean dimension d , the exponent maximizing efficiency has the minimum value $\eta_{\text{eff}} = (d + 1)/d$. At the opposite limit, the least efficient configuration (corresponding to a space-filling spiral or chain) yields the maximum value $\eta = 2$. This means that both river ($d = 2$) and vascular ($d = 3$) networks display an optimized degree of efficiency and suggests that their evolution shaped them to minimize the cost functions C_i . Our results show that the value of the exponent is smaller in food webs than in other systems. This is due to the absence of any euclidean dimension d (species are not constrained

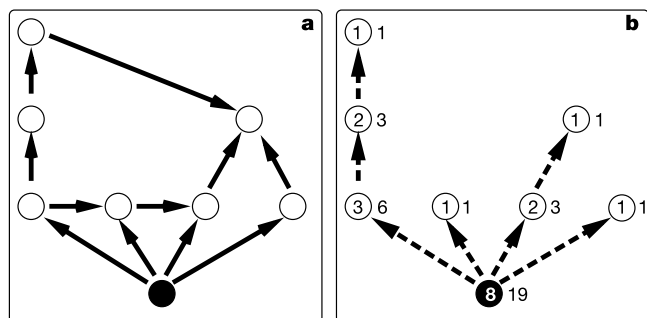


Figure 1 Computation of A_i and C_i from a spanning tree of the food web. **a**, Example of a food web with $S = 7$. The black vertex is the external environment. Species are layered in trophic levels. **b**, A possible spanning tree of the web. A_i is in the centre of each vertex i , C_i on the right. (See the text and Methods for details.)

to occupy the 'sites' of a regular lattice on a plane or in the space). In other words, spanning trees derived from food webs can in principle range between two extreme topologies: the chain-like (the least efficient, with all species feeding sequentially on one another, yielding $\eta = 2$ as before) and the star-like (the 'new' optimized state, with all species feeding directly on the environment, corresponding to the limit $\eta \rightarrow 1$) configuration. Remarkably, food web topology is more efficient than that of rivers or vascular systems; however, the optimal star-like configuration is not realized. This is related to the number of trophic levels¹ (the number of links in the shortest chains from the environment to the species; see Methods) in a food web: a tree spanning S species is chain-like if there are S distinct trophic levels, and star-like if there is only one level (which yields $C_i = A_i = 1$ for all $i = 1, S$ and $C_0 = 2A_0 - 1$, trivially corresponding to $\eta = 1$). As a result of competition, not every species can prey on the same resource. Species therefore tend to differentiate and occupy more than one trophic level to coexist²³. However, empirical observations^{1-3,6-15} reveal that real food webs display a small number of levels (usually in the range two to four) even if the number of species is large, which is consistent with our finding that η is closer to 1 than to 2.

We showed that the scaling exponent η is independent of other topological quantities, in particular the connectance $c = L/S^2$, which is the ratio of observed to possible links¹² (ranging from $c = 0.02$ in grassland to $c = 0.31$ in Skipwith Pond). Although a set of webs is consistent with the 'constant connectance' hypothesis²⁷ (according to which c is expected to be approximately constant about the value 0.1), two extreme deviations from this prediction are observed. The large values ($c = 0.31$) observed in Coachella Valley and Skipwith Pond are probably related to their small size²⁸. In other webs, c seems to decrease as the fraction of parasitic links (from larger to smaller species) increases⁹: the addition of parasitic interactions¹³ to the Ythan Estuary web¹¹ decreases c from 0.06 to

0.04. The simultaneous presence of predators, pathogens and parasites¹⁰ in Silwood Park yields the intermediate value $c = 0.03$. Finally, grassland (describing almost only flows from a larger to smaller species⁹) has the smallest value $c = 0.02$, thus being very close to a tree-like structure. These results are consistent with Cousins's conjecture²⁹, which states that "the passage of energy from large to small organisms involves problems so great that the parasite has to specialize on a single species or single family of species". Our results suggest that these evolutionary constraints do not affect the shape of the spanning trees. This differs radically from the predictions of current food web models^{3,23,24}, where all topological features (including the form of the spanning tree) are tuned by the same parameter determining the connectance. As shown in Table 1, the Cascade³ and Niche²⁴ (static) models fail to reproduce the values of c and η simultaneously for any real web. A dynamic model such as the Webworld²³⁻³⁰ model (which describes evolving species in terms of a time-dependent set of features determining their interactions) succeeds in reproducing both quantities in its 'evolved' asymptotic state (see Table 1 and legend), but only for the webs consistent with the 'constant connectance' hypothesis. This occurs because, in the model, a species displaying a set of features can in principle evolve to display any other feature²³. According to Cousins's conjecture, this possibility becomes unrealistic for parasitic interactions: being a specialized parasite of a certain host is likely to prevent specialization against other species.

Recent studies²⁸ relate the connectance to the stability of food webs. The robustness of real webs against successive removals of either a randomly chosen species or the most connected species increases monotonically²⁸ with c (which is thus a measure of the stability). In our framework, this means that the stability increases as the number of 'weak' links increases ('strong' links contribute only a negligible term $1/S$ to c). Our results then indicate that 'strong' and 'weak' links might have complementary roles: the

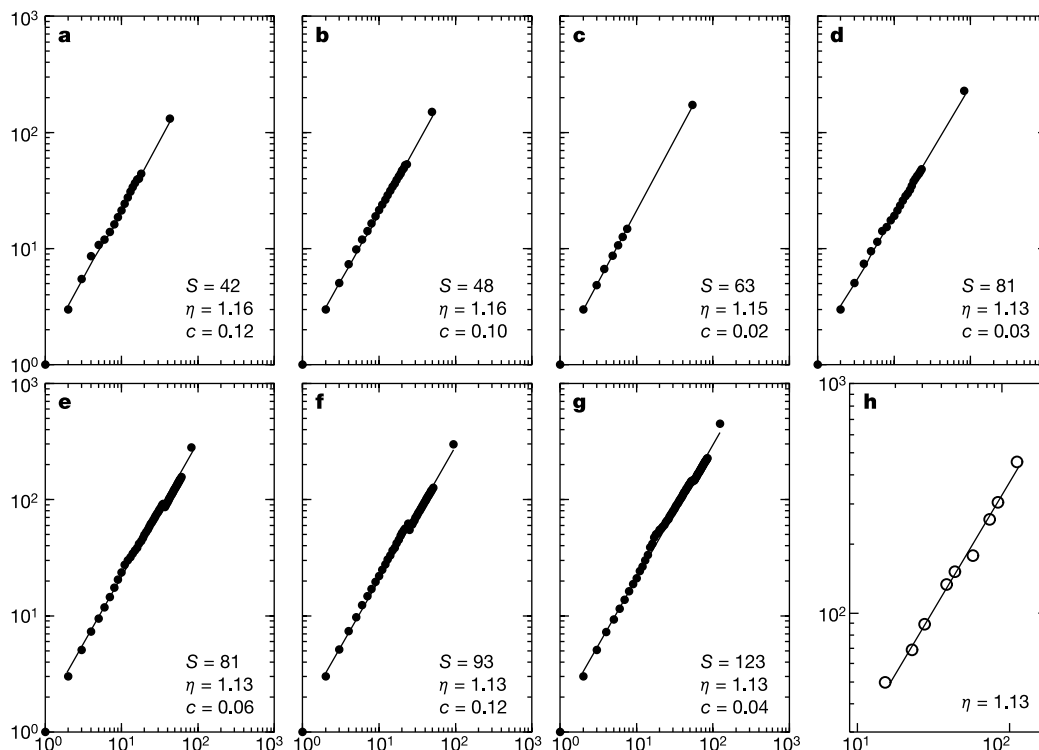


Figure 2 Scaling of C against A . Plot of scaling relations C_i against A_i (see the text and Methods) and the best power-law fit (the error in the value of the exponent η is always 0.03) to the data. **a**, St Martin Island⁷; **b**, St Marks Seagrass⁸; **c**, grassland⁹; **d**, Silwood Park¹⁰; **e**, Ythan Estuary without parasites¹¹; **f**, Little Rock Lake¹²; **g**, Ythan

Estuary with parasites¹³ (for each web, the number of species S and the connectance c are also reported). **h**, Plot of C_0 against A_0 for a collection of webs⁶⁻¹⁵ including those in **a-g**; see Methods (note that the scale is different).

Table 1 Predictions of food web models

C	η Cascade	η Niche	η^0 Webworld	η^∞ Webworld	C^∞ Webworld
0.005	1.12	1.03	1.83	1.61	0.08
0.01	1.09	1.04	1.51	1.20	0.10
0.05	1.08	1.06	1.33	1.11	0.12
0.1	1.06	1.08	1.22	1.09	0.18
0.3	1.04	1.07	1.18	1.07	0.37
0.5	1.01	1.04	1.05	1.04	0.50

The value of the exponent η (the error is always about 0.02) in different models is shown as a function of the connectance c . In static (Cascade³ and Niche²⁴) models the exponent (η Cascade or η Niche) is determined by a single tuning parameter related to c . In the Webworld model²³, a competition parameter tunes the initial state (characterized by the initial values c and η^0 Webworld). Then the model evolves and reaches an 'asymptotic' state characterized by a smaller exponent η^∞ Webworld and by a larger connectance C^∞ Webworld (both efficiency and stability increase; see the text). All webs have 1,000 species initially.

former determine the tree structure related to the efficiency of the webs, and the latter form the loops involved in network robustness. Whereas the different nature of the interactions affects the number of loops²⁹ (and hence the stability), the structure of the spanning trees seems to be invariant and universal. Correspondingly, although the stability properties are likely to be determined by different evolutionary processes, the efficiency of food webs might be the result of a common organizing principle. □

Methods

To define the minimal spanning tree, we discuss two different approaches. One possibility is considering the spanning tree defined by the collection of the shortest chains from the environment to every species (the number of links in these chains gives the trophic level^{1–3} of the species). If several spanning trees are still compatible with this prescription, we consider all of them and then perform an ensemble average (see below). Alternatively, we can select for each species the strongest chain from the environment; that is, the one delivering most resources to the species. We prefer to adopt the former definition to include a larger number of data sets in the analysis and to compare them with the models, which in most cases ignore link strength. However, we note that the chains obtained by the two criteria (minimizing chain lengths and maximizing chain strengths) are related. Generally, the transfer of resources along each link is 'inefficient' (only a small fraction of resources is transferred from prey to predator^{1,2}) and hence long chains are likely to be weaker than short ones¹. This argument is also consistent with the empirical observation² that only a fraction $\lambda = 0.1$ (known as the ecological efficiency) of the resources of the species in a trophic level is transferred to predators. However, we stress that our definition of 'weak' and 'strong' links does not necessarily reflect the actual link strength.

To obtain the spanning tree minimizing chain length, we first order the species in levels (labelled by $l = 0, 1, \dots, l_{\max}$). Conventionally, the environment belongs to level $l = 0$. All species preying on species at level l (but not at lower ones) belong to level $l + 1$. We then remove all the 'weak' links pointing from a species at a given level l_1 to a species at a level l_2 lower than or equal to l_1 . At this point, if each species is left with only one incoming link, the spanning tree is already obtained (all remaining links are 'strong') and the quantities A_i and C_i are directly computed and plotted. Otherwise, if some species still have more than one incoming link, we consider the ensemble of all possible spanning trees (each defined by a different set of 'strong' links) and compute the quantities A_i and C_i on each of them separately. Finally, for each value of A , the average of the corresponding values of C is computed and plotted as a function of A . This produces the curves shown in Fig. 2a–g. The point (A_0, C_0) is the same for every tree in the ensemble of possible ones. The point $(1, 1)$ always occurs as a finite-size effect (it corresponds to the 'leaves', where the tree stops branching) together with the point $(2, 3)$ independently of the large-scale behaviour of the scaling relations (see Fig. 1). A power-law exactly crossing both points would then miss all other points, unless the value of the exponent were 2. We therefore always exclude the point $(1, 1)$ from the fit to avoid forcing the exponent towards an artificially high value. We finally obtain the composite figure (Fig. 2h) by simply plotting together the points (A_0, C_0) of each of the previous panels plus the three additional webs described in the text (we recall that these points are independent of the specific spanning tree realization). The value of C_0 corresponding to $A_0 = 82$ is an average of the values of Silwood Park and Ythan Estuary without parasites (both webs have 81 trophic species).

Received 25 July 2002; accepted 20 March 2003; doi:10.1038/nature01604.

1. Lawton, J. H. in *Ecological Concepts* (ed. Cherret, J. M.) 43–78 (Blackwell Scientific, Oxford, 1989).
2. Pimm, S. L. *Food Webs* (Chapman & Hall, London, 1982).
3. Cohen, J. E., Briand, F. & Newman, C. M. *Community Food Webs: Data and Theory* Biomathematics 20 (Springer, Berlin, 1990).
4. Strogatz, S. H. Exploring complex networks. *Nature* **410**, 268–276 (2001).
5. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47–97 (2002).
6. Cohen, J. E. *Ecologists' Co-Operative Web Bank 1.00* (The Rockefeller Univ., New York, 1989).
7. Goldwasser, L. & Roughgarden, J. Construction and analysis of a large Caribbean food web. *Ecology* **74**, 1216–1233 (1993).
8. Christian, R. R. & Luczkovich, J. J. Organizing and understanding a winter's seagrass foodweb network through effective trophic levels. *Ecol. Mod.* **117**, 99–124 (1999).
9. Martinez, N. D., Hawkins, B. A., Dawah, H. A. & Feifarek, B. P. Effects of sampling effort on the characterization of food web structure. *Ecology* **80**, 1044–1055 (1999).

10. Memmott, J., Martinez, N. D. & Cohen, J. E. Predators, parasitoids and pathogens: species richness, trophic generality and body sizes in a natural web. *J. Anim. Ecol.* **69**, 1–15 (2000).
11. Hall, S. J. & Raffaelli, D. Food web patterns: lessons from a species-rich web. *J. Anim. Ecol.* **60**, 823–842 (1991).
12. Martinez, N. D. Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol. Monogr.* **61**, 367–392 (1991).
13. Huxham, M., Beany, S. & Raffaelli, D. Do parasites reduce the chances of triangulation in a real food web? *Oikos* **76**, 284–300 (1996).
14. Polis, G. A. Complex trophic interactions in deserts: an empirical critique of food web theory. *Am. Nat.* **138**, 123–155 (1991).
15. Warren, P. H. Spatial and temporal variation in the structure of a freshwater food web. *Oikos* **55**, 299–311 (1989).
16. Montoya, J. M. & Solé, R. V. Small world patterns in food webs. *J. Theor. Biol.* **214**, 405–412 (2002).
17. Williams, R. J., Berlow, E. L., Dunne, J. A. & Barabási, A.-L. Two degrees of separation in complex food webs. *Proc. Natl Acad. Sci. USA* **99**, 12913–12916 (2002).
18. Camacho, J., Guimerà, R. & Amaral, L. A. N. Robust patterns in food web structure. *Phys. Rev. Lett.* **88**, 228102 (2002).
19. Dunne, J. A., Williams, R. J. & Martinez, N. D. Food-web structure and network theory: the role of connectance and size. *Proc. Natl Acad. Sci. USA* **99**, 12917–12922 (2002).
20. Banavar, J. R., Maritan, A. & Rinaldo, A. Size and form in efficient transportation networks. *Nature* **399**, 130–132 (1999).
21. West, G. B., Brown, J. H. & Enquist, B. J. A general model for the origin of allometric scaling laws in biology. *Science* **276**, 122–126 (1997).
22. West, G. B., Brown, J. H. & Enquist, B. J. The fourth dimension of life: fractal geometry and allometric scaling of organisms. *Science* **284**, 1677–1679 (1999).
23. Caldarelli, G., Higgs, P. G. & McKane, A. J. Modelling coevolution in multispecies communities. *J. Theor. Biol.* **193**, 345–358 (1998).
24. Williams, R. J. & Martinez, N. D. Simple rules yield complex food webs. *Nature* **404**, 180–183 (2000).
25. Rodriguez-Iturbe, I. & Rinaldo, A. *Fractal River Basins: Chance and Self-Organization* (Cambridge Univ. Press, Cambridge, 1996).
26. McMahon, T. A. & Bonner, J. T. *On Size and Life* (Scientific American Library, New York, 1983).
27. Martinez, N. D. Constant connectance in community food webs. *Am. Nat.* **139**, 1208–1218 (1992).
28. Dunne, J. A., Williams, R. J. & Martinez, N. D. Network structure and biodiversity loss in food webs: robustness increases with connectance. *Ecol. Lett.* **5**, 558–567 (2002).
29. Cousins, S. H. Species diversity measurements: choosing the right index. *Trends Ecol. Evol.* **6**, 190–192 (1991).
30. Drossel, B., Higgs, P. G. & McKane, A. J. The influence of predator–prey population dynamics on the long-term evolution of food web structure. *J. Theor. Biol.* **208**, 91–107 (2001).

Acknowledgements We acknowledge support from the FET Open Project IST-2001-33555 COSIN.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.C. (gcald@pil.phys.uniroma1.it).

Defective membrane repair in dysferlin-deficient muscular dystrophy

Dimple Bansal*, Katsuya Miyake†, Steven S. Vogel‡, Séverine Groh*, Chien-Chang Chen*, Roger Williamson§, Paul L. McNeil† & Kevin P. Campbell*

* Howard Hughes Medical Institute, Department of Physiology and Biophysics and Department of Neurology, University of Iowa Roy J. and Lucille A. Carver College of Medicine, Iowa City, Iowa 52242, USA
† Department of Cellular Biology and Anatomy, The Medical College of Georgia, Augusta, Georgia 30912, USA
‡ Laboratory of Molecular Physiology, National Institute of Alcohol Abuse and Alcoholism, National Institutes of Health, Rockville, Maryland 20852, USA
§ Department of Obstetrics and Gynecology, University of Iowa College of Medicine, Iowa City, Iowa 52242, USA

Muscular dystrophy includes a diverse group of inherited muscle diseases characterized by wasting and weakness of skeletal muscle¹. Mutations in dysferlin are linked to two clinically distinct muscle diseases, limb-girdle muscular dystrophy type 2B and Miyoshi myopathy, but the mechanism that leads to

muscle degeneration is unknown^{2,3}. Dysferlin is a homologue of the *Caenorhabditis elegans fer-1* gene, which mediates vesicle fusion to the plasma membrane in spermatids⁴. Here we show that dysferlin-null mice maintain a functional dystrophin-glycoprotein complex but nevertheless develop a progressive muscular dystrophy. In normal muscle, membrane patches enriched in dysferlin can be detected in response to sarcolemma injuries. In contrast, there are sub-sarcolemmal accumulations of vesicles in dysferlin-null muscle. Membrane repair assays with a two-photon laser-scanning microscope demonstrated that wild-type muscle fibres efficiently reseal their sarcolemma in the presence of Ca^{2+} . Interestingly, dysferlin-deficient muscle fibres are defective in Ca^{2+} -dependent sarcolemma resealing. Membrane repair is therefore an active process in skeletal muscle fibres, and dysferlin has an essential role in this process. Our findings show that disruption of the muscle membrane repair machinery is responsible for dysferlin-deficient muscle degeneration, and highlight the importance of this basic cellular mechanism of membrane resealing in human disease.

To explore the function of dysferlin in skeletal muscle, dysferlin-null mice were generated by homologous recombination. The targeting vector was designed to carry an approximately 12-kilobase deletion at the 3' end of the gene, deleting the last three coding exons including the transmembrane domain (Supplementary Fig. 1). The genotypes of the mice were verified by PCR (Fig. 1a). The complete loss of protein expression was confirmed by western blot analysis of skeletal muscle microsomes (Fig. 1b). Immunofluorescence analysis revealed localization of dysferlin in both sarcolemma and cytoplasmic vesicles in wild-type muscle as reported previously⁵, but no staining was observed in dysferlin-null muscle (Fig. 1c, top row). Dysferlin has been associated with calpain 3 and caveolin-3, mutations of which are responsible for limb-girdle muscular dystrophy type 2A (LGMD2A) and LGMD1C, respectively¹, although the nature of this association remains unclear^{6,7}. However, no changes were observed in caveolin-3 (Fig. 1b, c) or calpain 3 (Fig. 1b) expression in dysferlin-null muscle.

Histological and other observations of the skeletal muscle revealed that dysferlin-null mice develop a slowly progressive muscular dystrophy (Fig. 2a). By the age of 2 months, a few individual necrotic and centrally nucleated fibres could be detected throughout the muscle; the number increased with age. By 8 months, the muscle had developed all the pathological characteristics of muscular dystrophy such as regenerating fibres, split fibres,

muscle necrosis with macrophage infiltration and fat replacement. Although all the examined muscles showed muscle pathology, the severity of the pathology varied in different muscles of dysferlin-null mice (Supplementary Fig. 2). Plasma membrane disruptions were identified with Evans blue, a membrane-impermeant dye. Unlike dystrophin-glycoprotein complex (DGC)-linked muscular dystrophies, in which clusters of Evans-blue-positive fibres are observed^{8,9}, dysferlin-null mice showed individual positive fibres, suggesting that the injury and the disease progression take place at the level of single muscle fibres (Fig. 2b). Consistent with the Evans blue analysis was the observation that serum creatine kinase levels of dysferlin-null mice ($1,022 \pm 368 \text{ U l}^{-1}$, mean $\pm$ s.d., $n = 8$) were severalfold higher than those of wild-type mice ($157 \pm 84 \text{ U l}^{-1}$, $n = 5$). Muscle regeneration is a response to muscle degeneration commonly seen in muscular dystrophy. A muscle undergoing active regeneration has centrally nucleated muscle fibres and large variability in fibre size. About 10% of all fibres in dysferlin-null muscle were centrally nucleated by 2 months of age, 20% by 4 months and 48% and 65% by 8 and 10 months, respectively (Fig. 2c). Centrally

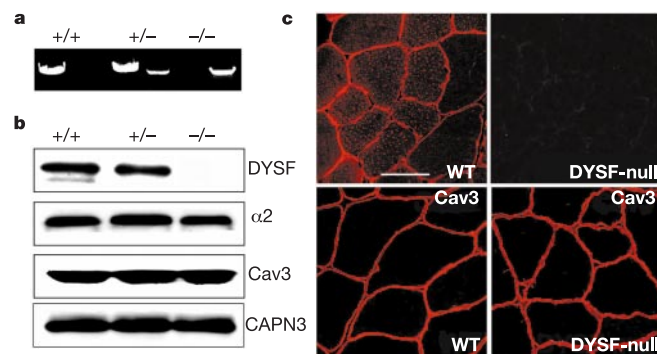


Figure 1 Complete loss of dysferlin expression in dysferlin-null mice. **a**, PCR genotyping of the wild-type (+/+), heterozygous (+/-) and dysferlin-null (-/-) mice. **b**, Loss of dysferlin (DYSF) expression by western blot analysis. Expression analysis of caveolin-3 (Cav3) and calpain 3 (CAPN3) and $\alpha 2$ (calcium channel subunit, for equal loading control). **c**, Immunofluorescence staining for dysferlin on skeletal muscle shows sarcolemma and vesicular cytoplasmic staining in wild-type muscle (WT) but a complete loss of staining in dysferlin-null (DYSF-null) muscle. Lower panels, normal staining for caveolin-3 in WT and DYSF-null skeletal muscle. Scale bar, 50 μm .

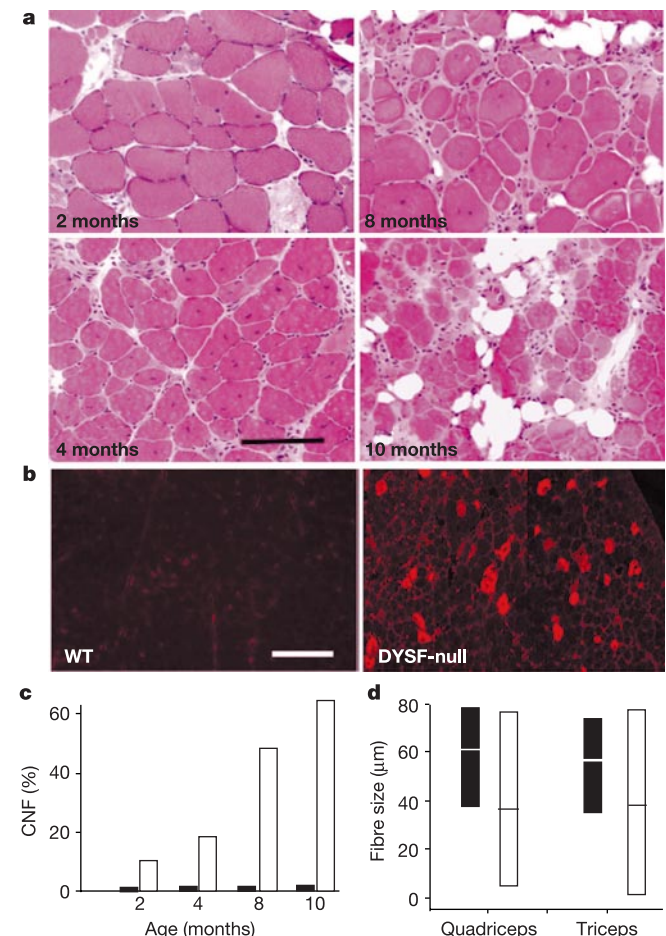


Figure 2 Dysferlin-null mice develop progressive muscular dystrophy. **a**, Haematoxylin-eosin-stained dysferlin-null skeletal muscle (quadriceps muscle) sections at 2, 4, 8 and 10 months of age. Scale bar, 100 μm . **b**, Individual Evans-blue-positive fibres can be detected in dysferlin-null (DYSF-null) muscle (9-month-old mice) but not in wild-type muscle (WT). Scale bar, 150 μm . **c**, Percentage of centrally nucleated fibres (CNF) increases with age in dysferlin-null muscle (open bars) but not in WT muscle (filled bars). For each sample set, $n > 1,000$; the data shown are an average of three sample sets. **d**, Greater fibre size variation was observed in dysferlin-null muscle (open bars) than in WT muscle (filled bars). The central horizontal line in each bar is the fibre diameter mean, and the size of each box is the size variation; $n = 3$ or 4 muscles from 9-month-old mice, with a total of 500–700 fibres.

nucleated fibres in wild-type mice never exceeded 2%. A much greater fibre size variability was also observed in dysferlin-null skeletal muscle than in wild-type muscle (Fig. 2d).

Several types of muscular dystrophy are linked to components of the DGC^{1,10}. In the DGC-linked muscular dystrophies, mutation in one of the DGC components disrupts the stability of the whole complex at the sarcolemma^{9,11}, which in turn disrupts the structural link between the sub-sarcolemma cytoskeleton and the extracellular matrix leading to an unstable sarcolemma. Disruption of this link makes the muscle more susceptible to contraction-induced injuries^{1,12,13}. To test whether the loss of dysferlin has any effect on the stability of DGC components, the expression of various DGC components was evaluated by immunofluorescence (Fig. 3a) and western blot analysis on KCl-washed skeletal muscle membranes (Fig. 3b). However, no change was observed in the expression of the DGC components in dysferlin-null muscle. Furthermore, biochemical purification of the DGC from skeletal muscle also suggests

the presence of a stable and functional DGC in dysferlin-null mice (Supplementary Fig. 3). We further tested the structural stability of the sarcolemma by exercising wild-type, dysferlin-null and mdx mice downhill on a treadmill¹⁴ (-10° angle) for 30 min at a speed of 25 m min^{-1} . A comparison of Evans blue uptake and serum creatine kinase levels, before and after running, was used as a measure for the sarcolemma damage. The mdx mice have a mutation in the *dystrophin* gene¹⁰, which causes disruption of the DGC and an unstable sarcolemma. This form of exercise, as expected, caused extensive membrane damage in the muscle of mdx mice, reflected by the significant increase in the serum creatine kinase levels and Evans blue uptake after exercise. By contrast, minimal damage was observed in wild-type and dysferlin-null mice, suggesting the presence of a stable sarcolemma in dysferlin-null mice (Fig. 3c). Taken together, these results suggest that the dysferlin-null mice have a functional DGC and that the mechanism of muscle degeneration in LGMD2B and Miyoshi myopathy patients is different from the DGC-linked muscular dystrophies.

Mechanical stress can cause muscle fibre necrosis by two separate mechanisms. First, as already mentioned, the sarcolemma of a muscle fibre could be more susceptible to injury, as in DGC-linked muscular dystrophies. Second, a muscle fibre could have a defective sarcolemma repair mechanism. Because the sarcolemma is structurally stable in dysferlin-null muscle, we propose that dysferlin has a role in the sarcolemma repair process. Membrane repair is a basic cellular mechanism that is required to reseal plasma membrane disruptions. Because plasma membrane disruption is a physiological event in skeletal muscle^{13,15}, the result of defective repair would be muscle fibre degeneration. Previous studies showed that membrane repair requires the accumulation and fusion of vesicles near the site of membrane disruptions¹⁶⁻¹⁹. Electron microscopic analysis of dysferlin-null muscle revealed sites of sarcolemma disruption associated with underlying accumulations of vesicles (Fig. 4a).

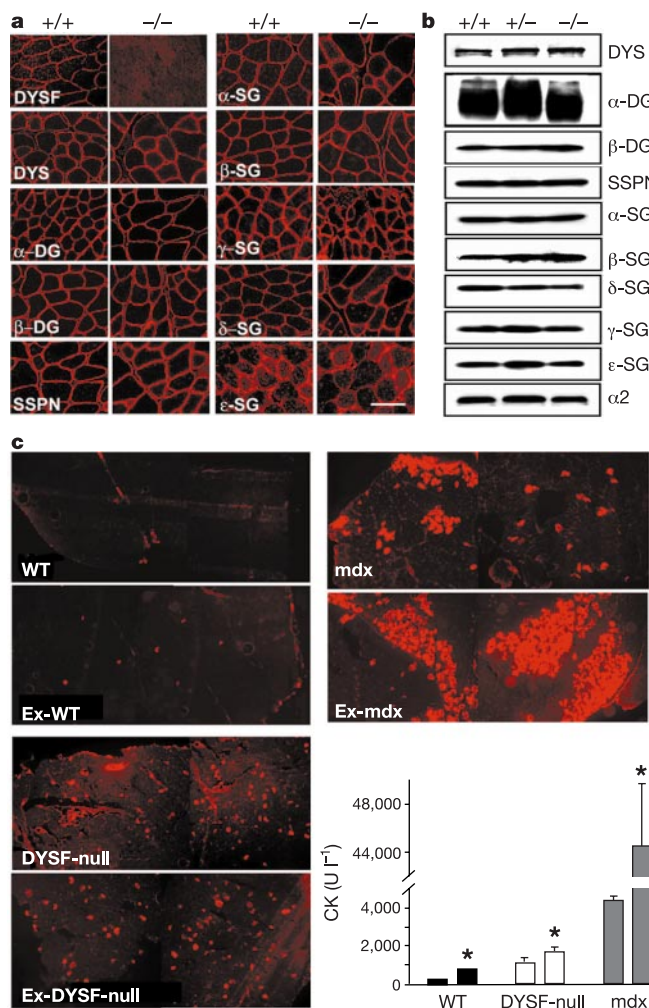


Figure 3 Normal expression of DGC components and a structurally stable sarcolemma in dysferlin-null mice. **a**, No staining for dysferlin (DYSF) but normal immunofluorescence of dystrophin (DYS), α -dystroglycan (α -DG), β -dystroglycan (β -DG), α -sarcoglycan (α -SG), β -sarcoglycan (β -SG), δ -sarcoglycan (δ -SG), γ -sarcoglycan (γ -SG), ϵ -sarcoglycan (ϵ -SG) and sarcospan (SSPN) in dysferlin-null muscle. Scale bar, 100 μm . **b**, Western blotting shows normal expression of DGC components in dysferlin-null mice; $\alpha 2$ used as a loading control. **c**, Significant increase in Evans blue uptake and serum creatine kinase (CK) levels was observed after exercise in mdx mice but not in wild-type (WT) and dysferlin-null (DYSF-null) mice ($n = 5$). Vertical bars (asterisks) are creatine kinase levels after exercise. Data are means $\pm$ s.d.

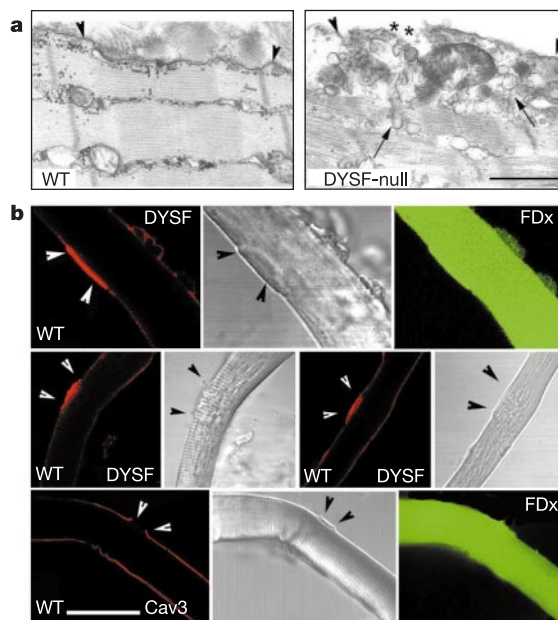


Figure 4 Vesicle accumulation and dysferlin-enriched membrane patches on the damaged muscle fibres. **a**, Electron micrograph of wild-type (WT) and dysferlin-null (DYSF-null) skeletal muscle. Sites with disrupted sarcolemma (asterisks) and vesicle accumulation (arrows) below the membrane (arrowheads) could be detected only in dysferlin-null muscle. Scale bar, 1 μm . **b**, Top two rows, membrane patches and wounds (bright field, arrowhead) enriched in dysferlin could be detected on the WT damaged fibres (positive for fluorescein dextran (FDx)). Bottom row, loss of caveolin-3 (Cav3) staining at the membrane patch sites was observed. Scale bar, 50 μm .

Similar morphologies were not detected in wild-type muscle. This result and several other observations, such as the homology of dysferlin to FER-1 in *C. elegans* (which has a role in vesicle fusion to the plasma membrane), the presence of six C2 domains in dysferlin, and a protein structure similar to that of synaptotagmins, are consistent with our hypothesis that dysferlin functions in plasma membrane repair. Synaptotagmins are a family of proteins with a role in vesicle trafficking and fusion to the plasma membrane, and regulate important cellular events including membrane repair in certain cell types²⁰. Moreover, the first C2 domain of dysferlin binds to phospholipids in a Ca^{2+} -dependent manner²¹.

To evaluate the role of dysferlin in membrane repair, dysferlin distribution was analysed in damaged single muscle fibres isolated from wild-type mice. Muscle fibres were damaged in the presence of a membrane-impermeant fluorescein dextran (FDx) to mark those fibres that experienced membrane damage. Dysferlin distribution was analysed by immunofluorescence on permeabilized fibres. The undamaged fibres (FDx-negative) showed intact sarcolemma and uniform membrane staining of dysferlin (not shown). Membrane wounds and patches of various sizes were detected (Fig. 4b, bright field) on the damaged fibres (FDx-positive). Interestingly, immunofluorescence analysis revealed the enrichment of dysferlin in these patches (Fig. 4b, arrowheads in top two rows). According to the 'patch hypothesis'²², membrane repair requires the accumulation and fusion of vesicle populations with each other and with the plasma membrane at or near the disruption. A 'patch' is thereby added across a defect, resealing the disrupted membrane¹⁶. Proteins with a direct role in this process can be detected at the disruption sites^{17,20}. We suggest that these dysferlin-enriched membrane patches correspond to the membrane wound sites on the damaged muscle fibres. To confirm this, we stained damaged fibres with the known sarcolemma proteins such as caveolin-3 (Fig. 4b, bottom row) and δ -sarcoglycan^{1,14} (not shown). A membrane disruption would cause a transient loss and reduction of the sarcolemma proteins at the wound sites. We detected a range from reduced staining to complete loss of these sarcolemma markers on the patch sites, confirming that these are indeed the wound sites on the sarcolemma. Our demonstration of dysferlin enrichment on the membrane disruption sites therefore strongly suggests a direct role of dysferlin in the membrane repair process.

To explore this potential mechanism further, a membrane repair assay was used²³ in a direct assessment of the resealing efficiency of dysferlin-null muscle. This assay was performed on single muscle fibres isolated from the flexor digitorum brevis muscle of the dysferlin-null and wild-type mice. Muscle fibres were irradiated with the mode-locked laser of a two-photon confocal scanning microscope to create plasma membrane disruptions. This injury was performed in the presence of FM 1-43, a membrane-impermeant fluorescent dye. FM 1-43 is soluble in both water and lipid, does not cross intact bilayers, and fluoresces only in lipidic environments such as cell membranes. At steady state, therefore, the membrane fluorescence of a wounded, unsealed muscle fibre will increase many-fold owing to the presence of endo-membrane that becomes stained with the dye. Resealing of the membrane halts this increase in fluorescence. The fluorescence that developed near the disruption (laser irradiation) site was imaged and measured at 10-s intervals beginning 20 s before ($t = 0$) and extending until about 3 min after the damage was induced. The initial fluorescence intensity (before the damage was induced) was set to 0 and the values show the increase in fluorescence intensity after the damage was induced. Wild-type muscle fibres, damaged in the presence of Ca^{2+} , significantly impeded further dye entry at about the 1-min time point after disruption (Fig. 5a, uppermost row of panels, and Fig. 5b, filled circles), suggesting that, on average, these fibres resealed within 1 min in the presence of Ca^{2+} . In contrast, wild-type fibres damaged in the absence of Ca^{2+} displayed comparatively little ability to impede dye entry over the entire time course of the

experiment, suggesting a failure of resealing (Fig. 5a, second row of panels, and Fig. 5c, filled circles). This result suggests that muscle fibres can reseal their membrane efficiently and that membrane resealing is a Ca^{2+} -dependent process in skeletal muscle similar to that in many other cell types. Interestingly, dysferlin-null fibres in the presence of Ca^{2+} failed to halt dye uptake after an identical laser-induced disruption was created (Fig. 5a, third row of panels, and Fig. 5b, open triangles). Indeed, the apparently unimpeded uptake of FM 1-43 observed in dysferlin-null fibres in the presence of Ca^{2+} was similar to that observed in the wild-type and dysferlin-null (Fig. 5a, bottom row of panels, and Fig. 5c, open triangles) experiments conducted in the absence of resealing-permissive Ca^{2+} . Unlike dysferlin-null fibres, membrane repair assays conducted on muscle fibres in mdx mice showed efficient membrane resealing in the presence of Ca^{2+} similar to that in wild-type muscle (P. McNeil, personal communication). These results indicate that dysferlin-null muscle fibres are defective in Ca^{2+} -dependent resealing of sarcolemma disruptions.

We suggest that dysferlin has a role during the membrane fusion step of the repair process in muscle fibres. Such a role is consistent with the presence of several C2 domains in dysferlin and with its homology to several known proteins with a role in vesicle fusion. We propose that dysferlin, present on the vesicles, facilitates vesicle docking and fusion with the plasma membrane either by interacting

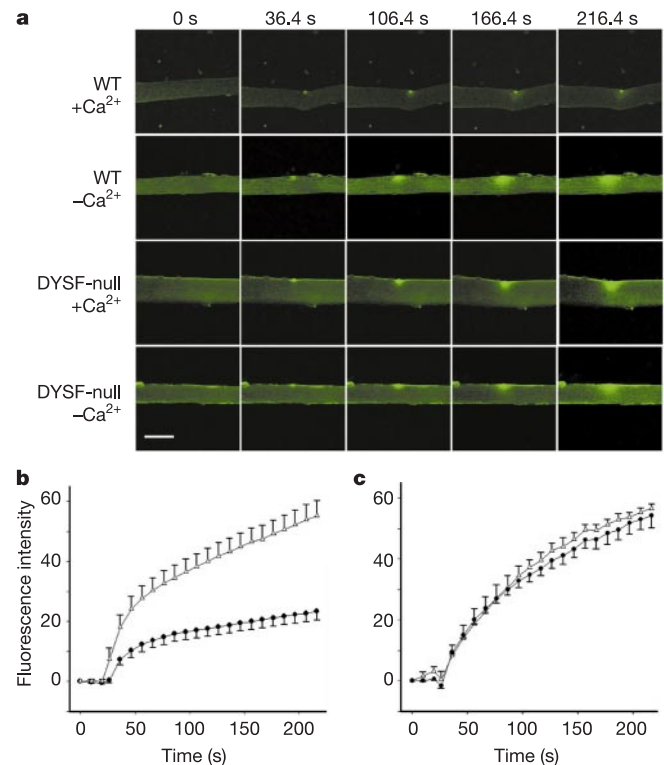


Figure 5 Defective resealing of membranes in dysferlin-null muscle. **a**, Membrane repair assay performed on wild-type (WT) and dysferlin-null (DYSF-null) single muscle fibres in the presence and absence of Ca^{2+} . The damage was induced at $t = 20$ s. The top two rows of panels show the assay performed on wild-type fibres in the presence (WT + Ca^{2+}) and absence (WT - Ca^{2+}) of Ca^{2+} . The bottom two rows of panels show the repair assay performed on dysferlin-null fibres in the presence (DYSF-null + Ca^{2+}) and absence (DYSF-null - Ca^{2+}) of Ca^{2+} . Scale bar, 50 μm . **b**, Plot of fluorescence intensity ($n = 16$) against time in WT (filled circles) and dysferlin-null (open triangles) fibres in the presence of Ca^{2+} . Data are means $\pm$ s.e. **c**, Measurement of fluorescence intensity ($n = 4$) versus time in WT (filled circles) and dysferlin-null (open triangles) fibres in the absence of Ca^{2+} . Data are means $\pm$ s.e.; statistical analysis was done with Student's *t*-test. $P < 0.05$ is considered significant.

with the other dysferlin molecules or with some other unknown protein-binding partner(s) at the plasma membrane.

The finding that dysferlin-null muscle is inefficient in plasma membrane resealing highlights the importance of this basic cellular function in various forms of muscular dystrophy. Because cell membrane repair is required for surviving a mechanical injury, and given that cell wounding is common in many mechanically active mammalian cells, defects in this repair pathway may be linked to other human diseases. This study reveals a novel mechanism of muscular dystrophy in which the membrane repair machinery of muscle is defective, as opposed to several other types of muscular dystrophies in which the primary defect is in the stability of sarcolemma. Dysferlin is the first identified member of the membrane repair machinery in skeletal muscle. This finding will help to identify other members of the muscle membrane repair machinery, and those will serve as potential candidate genes for other muscular dystrophies and muscle-related diseases with unknown aetiology. □

Methods

Generation of dysferlin-null mice

The targeting vector was constructed to delete the last three coding exons including the exons coding for the transmembrane domain of the dysferlin gene. The targeting vector was electroporated into R1 embryonic stem cells, and clones carrying the specific targeted locus were confirmed by PCR screening and sequencing of the PCR product. Targeted clones were used to generate chimaeric mice by blastocyst injections. Heterozygous mice obtained from the chimaeric mice were bred together to obtain homozygous mutant mice. Homozygous mutant mice were backcrossed with C57/BL6 wild-type mice for two generations. Wild-type littermates were used as controls. Dysferlin-null mice are viable and have a normal growth rate.

Membrane repair assay

This experiment was performed on four wild-type and four dysferlin-null mice at 6 weeks of age, blind to the genotype. For every mouse, the assay was performed on four different fibres in the presence of Ca^{2+} and one or two fibres in the absence of Ca^{2+} . Single muscle fibres were isolated from the flexor digitorum brevis muscle by treatment with collagenase I. Fibres were washed and resuspended in Dulbecco's PBS containing 1 mM Ca^{2+} (Gibco) for the resealing experiment in the presence of Ca^{2+} , and PBS containing 1 mM EGTA for the resealing experiment in the absence of Ca^{2+} . Single fibres were mounted on a glass slide chamber and membrane damage was induced in the presence of FM 1-43 dye (2.5 μM ; Molecular Probes) with a two-photon confocal laser-scanning microscope (LSM 510; Zeiss) coupled to a 10-W Argon/Ti:sapphire laser²³. To induce damage, a 5 $\mu\text{m} \times 5 \mu\text{m}$ area of the sarcolemma on the surface of the muscle fibre was irradiated at full power for 6.4 s at $t = 20$ s. Images were captured beginning 20 s before ($t = 0$) and for 3 min after the irradiation at 10-s intervals. For every image taken, the fluorescence intensity at the site of the damage was measured with the Zeiss LSM 510 imaging software. Fibres that had no membrane resealing showed dye filling at the wound site over the entire course of the experiment, whereas dye influx halted within 1 min for the fibres that resealed under the experimental conditions.

Analysis of dysferlin expression on damaged muscle fibres

Single muscle fibres were damaged by being flushed through an 18-gauge syringe in the presence of lysine-fixable fluorescein dextran dye (10 kDa, 10 mg ml^{-1} in PBS; Molecular Probes). After incubation for 2 min at 37 °C, fibres were transferred to ice, washed with chilled PBS and fixed in 4% paraformaldehyde for 10 min. Fibres were then washed with PBS and permeabilized with detergent; immunofluorescence analysis was performed with antibodies against dysferlin, caveolin-3 and δ -sarcoglycan. Dysferlin, caveolin-3 and δ -sarcoglycan distribution was analysed on more than 200 damaged muscle fibres: 25 fibres showed dysferlin-enriched sites, but none of them showed any kind of membrane enrichment for caveolin-3 or δ -sarcoglycan.

Biochemical, immunofluorescence and histological analysis

KCl-washed skeletal muscle membranes were prepared and western blot analysis was performed as described previously^{24,25}. DGC purification from the total muscle homogenates was performed as described previously²⁵. Western blot analysis of dysferlin was performed using an amino-terminal rabbit polyclonal antibody. Rabbit polyclonal antibody against mouse dysferlin was generated by injecting New Zealand white rabbits intramuscularly and subcutaneously with an N-terminal dysferlin peptide (LLADCSQPLCDIHEIPSATH; single-letter amino acid codes). Antibody specificity was verified by competition experiments with the corresponding peptide on the western blot. Antibodies against DGC components were used as described previously^{9,14}. Expression analysis of caveolin-3 and calpain 3 was performed with mouse monoclonal antibodies; caveolin-3 (Transduction Laboratories) and calpain 3 (Calp3c12A2; Novocastra). Immunofluorescence was performed on 7- μm transverse cryosections as described previously⁹. For dysferlin immunostaining, the monoclonal antibody Hamlet (Novocastra)²⁶ was used at 1:25 dilution with an *in situ* antigen denaturation method described previously²⁷. One-year-old dysferlin-null and wild-type mice were perfused with

2% paraformaldehyde in PBS and quadriceps muscle was dissected out for electron microscopic analysis. Tissue processing was done as described previously⁵. Histological analysis of skeletal muscle and Evans blue injections and exercise experiments were performed as described previously^{8,9,14}.

Received 27 December 2002; accepted 14 March 2003; doi:10.1038/nature01573.

- Cohn, R. D. & Campbell, K. P. Molecular basis of muscular dystrophies. *Muscle Nerve* **23**, 1456–1471 (2000).
- Bashir, R. *et al.* A gene related to *Caenorhabditis elegans* spermatogenesis factor *fer-1* is mutated in limb-girdle muscular dystrophy type 2B. *Nature Genet.* **20**, 37–42 (1998).
- Liu, J. *et al.* Dysferlin, a novel skeletal muscle gene, is mutated in Miyoshi myopathy and limb girdle muscular dystrophy. *Nature Genet.* **20**, 31–36 (1998).
- Achanzar, W. E. & Ward, S. A nematode gene required for sperm vesicle fusion. *J. Cell Sci.* **110**, 1073–1081 (1997).
- Piccolo, E., Moore, S. A., Ford, G. C. & Campbell, K. P. Intracellular accumulation and reduced sarcolemmal expression of dysferlin in limb-girdle muscular dystrophies. *Ann. Neurol.* **48**, 902–912 (2000).
- Matsuda, C. *et al.* The sarcolemmal proteins dysferlin and caveolin-3 interact in skeletal muscle. *Hum. Mol. Genet.* **10**, 1761–1766 (2001).
- Anderson, L. V. *et al.* Secondary reduction in calpain 3 expression in patients with limb girdle muscular dystrophy type 2B and Miyoshi myopathy (primary dysferlinopathies). *Neuromuscul. Disord.* **10**, 553–559 (2000).
- Straub, V., Rafael, J. A., Chamberlain, J. S. & Campbell, K. P. Animal models for muscular dystrophy show different patterns of sarcolemmal disruption. *J. Cell Biol.* **139**, 375–385 (1997).
- Duclos, F. *et al.* Progressive muscular dystrophy in α -sarcoglycan-deficient mice. *J. Cell Biol.* **142**, 1461–1471 (1998).
- Durbeej, M. & Campbell, K. P. Muscular dystrophies involving the dystrophin-glycoprotein complex: an overview of current mouse models. *Curr. Opin. Genet. Dev.* **12**, 349–361 (2002).
- Ervasti, J. M., Ohlendieck, K., Kahl, S. D., Gaver, M. G. & Campbell, K. P. Deficiency of a glycoprotein component of the dystrophin complex in dystrophic muscle. *Nature* **345**, 315–319 (1990).
- Petrof, B. J., Shrager, J. B., Stedman, H. H., Kelly, A. M. & Sweeney, H. L. Dystrophin protects the sarcolemma from stresses developed during muscle contraction. *Proc. Natl Acad. Sci. USA* **90**, 3710–3714 (1993).
- Clarke, M. S., Khakee, R. & McNeil, P. L. Loss of cytoplasmic basic fibroblast growth factor from physiologically wounded myofibers of normal and dystrophic muscle. *J. Cell Sci.* **106**, 121–133 (1993).
- Coral-Vazquez, R. *et al.* Disruption of the sarcoglycan-sarcospan complex in vascular smooth muscle: a novel mechanism for cardiomyopathy and muscular dystrophy. *Cell* **98**, 465–474 (1999).
- McNeil, P. L. & Khakee, R. Disruptions of muscle fiber plasma membranes. Role in exercise-induced damage. *Am. J. Pathol.* **140**, 1097–1109 (1992).
- McNeil, P. L. & Terasaki, M. Coping with the inevitable: how cells repair a torn surface membrane. *Nature Cell Biol.* **3**, E124–E129 (2001).
- Bi, G. Q., Alderton, J. M. & Steinhardt, R. A. Calcium-regulated exocytosis is required for cell membrane resealing. *J. Cell Biol.* **131**, 1747–1758 (1995).
- Miyake, K. & McNeil, P. L. Vesicle accumulation and exocytosis at sites of plasma membrane disruption. *J. Cell Biol.* **131**, 1737–1745 (1995).
- Steinhardt, R. A., Bi, G. Q. & Alderton, J. M. Cell membrane resealing by a vesicular mechanism similar to neurotransmitter release. *Science* **263**, 390–393 (1994).
- Reddy, A., Caler, E. V. & Andrews, N. W. Plasma membrane repair is mediated by Ca^{2+} -regulated exocytosis of lysosomes. *Cell* **106**, 157–169 (2001).
- Davis, D. B., Doherty, K. R., Delmonte, A. J. & McNally, E. M. Calcium-sensitive phospholipid binding properties of normal and mutant ferlin C2 domains. *J. Biol. Chem.* **277**, 22883–22888 (2002).
- McNeil, P. L., Vogel, S. S., Miyake, K. & Terasaki, M. Patching plasma membrane disruptions with cytoplasmic membrane. *J. Cell Sci.* **113**, 1891–1902 (2000).
- McNeil, P. L., Miyake, K. & Vogel, S. S. The self-sealing membrane concept. *Proc. Natl Acad. Sci. USA* **100**, 4592–4597 (2003).
- Ohlendieck, K. & Campbell, K. P. Dystrophin-associated proteins are greatly reduced in skeletal muscle from mdx mice. *J. Cell Biol.* **115**, 1685–1694 (1991).
- Durbeej, M. *et al.* Disruption of the β -sarcoglycan gene reveals pathogenetic complexity of limb-girdle muscular dystrophy type 2E. *Mol. Cell* **5**, 141–151 (2000).
- Anderson, L. V. *et al.* Dysferlin is a plasma membrane protein and is expressed early in human development. *Hum. Mol. Genet.* **8**, 855–861 (1999).
- Miner, J. H. *et al.* The laminin alpha chains: expression, developmental transitions, and chromosomal locations of $\alpha 1$ –5, identification of heterotrimeric laminins 8–11, and cloning of a novel $\alpha 3$ isoform. *J. Cell Biol.* **137**, 685–701 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank all members of the Campbell laboratory for the critical reading of this manuscript, discussions and supplying critical reagents; and M. Hassebrock, S. Lowen, E. Hurst and K. Garringer for technical assistance. D. Bansal was supported by American Heart Predoctoral Fellowship (Heartland). We thank the University of Iowa DNA Core Facility, which is supported in part by the Diabetes Endocrinology Research Center and the University of Iowa Roy J. and Lucille A. Carver College of Medicine. We also thank University of Iowa Central Microscopy Research Facility and the Medical College of Georgia Imaging Core. This work was supported by the Muscular Dystrophy Association (K.P.C.). P.L.McN. is supported by NASA, and S.S.V. by NIH. K.P.C. is an investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to K.P.C. (kevin-campbell@uiowa.edu).

Identification of *Vangl2* and *Scrb1* as planar polarity genes in mammals

Mireille Montcouquiol^{*†}, Rivka A. Rachel^{†‡}, Pamela J. Lanford[§], Neal G. Copeland[‡], Nancy A. Jenkins[‡] & Matthew W. Kelley^{*}

^{*} Section on Developmental Neuroscience, NIDCD, National Institutes of Health, Rockville, Maryland 20850, USA

[‡] Mouse Cancer Genetics Program, National Cancer Institute-Frederick, Frederick, Maryland 21702, USA

[§] Biological Sciences, University of Maryland, College Park, Maryland 20742, USA

[†] These authors contributed equally to this work

In mammals, an example of planar cell polarity (PCP) is the uniform orientation of the hair cell stereociliary bundles within the cochlea. The PCP pathway of *Drosophila*^{1–4} refers to a conserved signalling pathway that regulates the coordinated orientation of cells or structures within the plane of an epithelium. Here we show that a mutation in *Vangl2*, a mammalian

homologue of the *Drosophila* PCP gene *Strabismus/Van Gogh*, results in significant disruptions in the polarization of stereociliary bundles in mouse cochlea as a result of defects in the direction of movement and/or anchoring of the kinocilium within each hair cell. Similar, but less severe, defects are observed in animals containing a mutation in the LAP protein family gene *Scrb1* (homologous with *Drosophila scribble*). Polarization defects in animals heterozygous for *Vangl2* and *Scrb1* are comparable with *Vangl2* homozygotes, demonstrating genetic interactions between these genes in the regulation of PCP in mammals. These results demonstrate a role for the PCP pathway in planar polarization in mammals, and identify *Scrb1* as a PCP gene.

Vangl2 (also called *Ltap* or *Lpp1*)—orthologous with *Drosophila Strabismus/Van Gogh* (*Stbm/Vang*)—encodes a four-transmembrane protein with a PDZ-domain binding motif at the carboxy terminus. Sequence analysis of vertebrate and invertebrate *Stbm/Vang* orthologues indicates that the gene has been highly conserved. Recently, two groups demonstrated that *Vangl2* is mutated in the loop-tail (*Lp*) mouse^{5,6}, a semi-dominant mutation that originated spontaneously in Strong's A strain⁷. Homozygous *Lp* embryos suffer from an open neural tube throughout the hindbrain and spinal

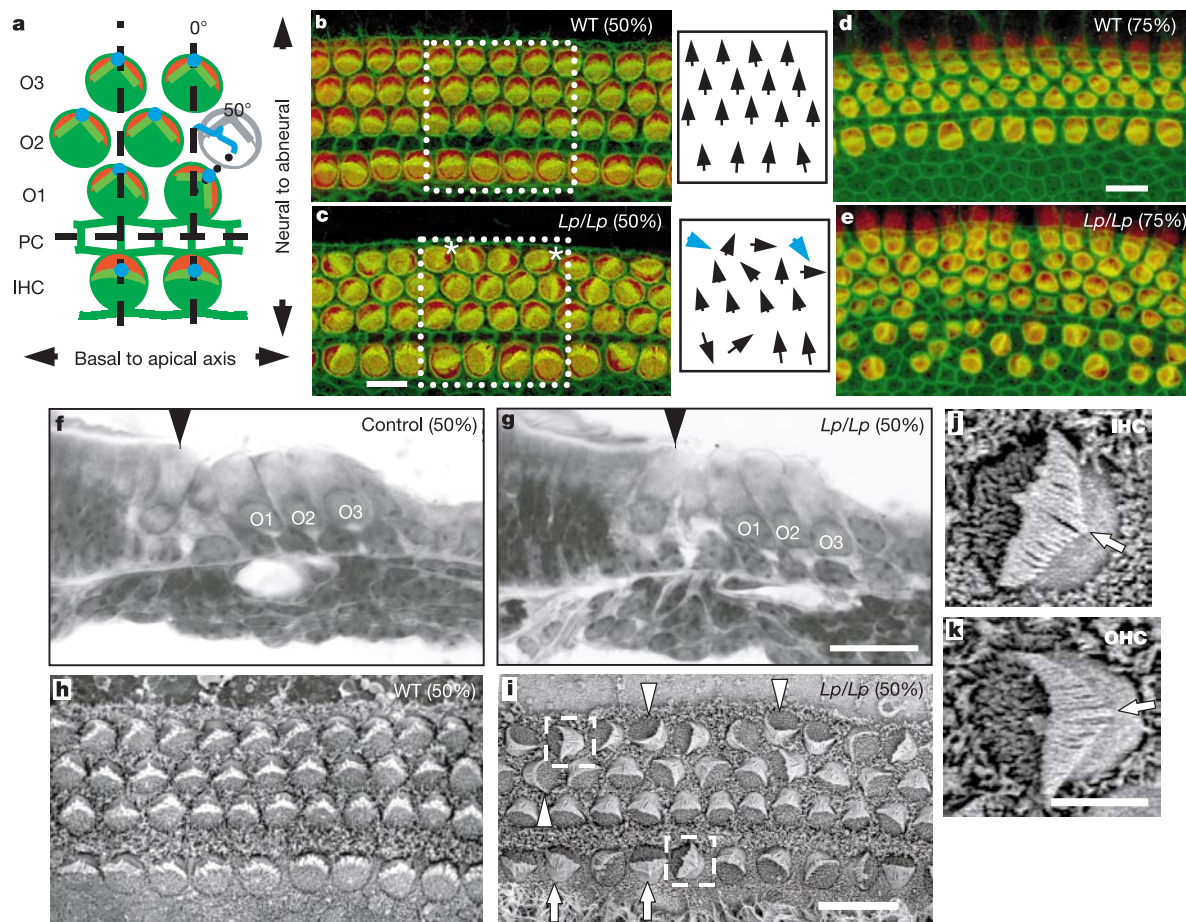


Figure 1 Stereociliary bundle orientation is disrupted in *Lp/Lp* cochleae. **a**, The stereociliary bundle (light green) on each IHC and OHC (O, numbered) is oriented such that the vertex is closest to the abneural edge of the epithelium. The method for determination of bundle orientation is shown on the right of O1. **b**, Luminal surface of the organ of Corti in the middle turn of an E18 wild-type cochlea. Stereociliary bundles (green) on IHCs and OHCs (both red) are uniformly oriented; orientations within the boxed region are indicated by arrows to the right. **c**, Middle turn in an E18 *Lp/Lp* cochlea. Bundle orientations are non-uniform with some bundles rotated by greater than 90° (asterisks and blue arrows). **d**, Apical turn in an E18 wild-type cochlea. By comparison with the base, development is

delayed; however, orientation is largely uniform. **e**, Apical turn in an E18 *Lp/Lp* cochlea. Note the marked disruption in bundle orientation. **f**, **g**, Cross-sections through the middle turn of control (**f**) and *Lp/Lp* (**g**) cochleae. A single IHC (arrow) and three OHCs (O numbered) are present in both, and cellular pattern appears comparable. **h**, **i**, SEM images of the middle turn in E18 wild-type (**h**) and *Lp/Lp* (**i**) cochlea. Bundle morphology appears normal in both; however, the orientations for many IHCs (arrows) and OHCs (arrowheads) are significantly disrupted in the *Lp/Lp* cochlea. **j**, **k**, High magnification images of an IHC (**j**) and OHC (**k**) from the cochlea in **i**. Scale bar: **b–e**, **h**, **i**, 10 μ m; **f**, **g**, 20 μ m; **j**, **k**, 3 μ m.

region, and typically die shortly before or at birth⁷.

One of the best examples of planar polarity in vertebrates is the uniform orientation of the stereociliary bundles located at the apex of each mechanosensory hair cell within the mammalian cochlea. Each bundle is comprised of a group of actin-based stereocilia arranged in a crescent shape with the vertex located on the side of the cell closest to the abneural edge of the epithelium (Fig. 1a). Hair cell transduction and auditory perception are dependent on the appropriate orientation of the stereociliary bundles⁸; however, the factors that mediate this orientation are unknown.

To determine whether *Vangl2* has a role in the generation of planar polarity in mammals, the orientation of stereociliary bundles was analysed in cochleae from *Lp/Lp* homozygotes and wild-type littermates at embryonic day 18 (E18). Significant disruptions in bundle orientation are present on both inner and outer hair cells (IHCs and OHCs, respectively) in cochleae from *Lp/Lp* mice (Fig. 1b–e). The degree of disruption varies from cell to cell, with some bundles rotated by as much as 180° from their normal orientation (Fig. 2a–d). Overall development of the cochlea, including the determination of cell fates and apical–basal polarity of hair cells, appears normal (Fig. 1f, g). Furthermore, analysis of bundle morphology by scanning electron microscopy (SEM) indicates that defects are limited to changes in orientation (Fig. 1h–k).

Cellular differentiation in the cochlea occurs in a gradient that initiates in the mid-basal turn and extends in both directions along the basal to apical axis⁹. In addition, differentiation in any given region occurs in a gradient that extends from neural (IHCs) to abneural (OHCs). To determine whether the relative stage of development might have a role in the degree of bundle mis-orientation, average deviations were determined on the basis of cell type and position along the basal-to-apical and neural-to-abneural axes. For OHCs, deviations within each row did not differ significantly along the basal-to-apical axis of the duct except for third row OHCs between the 25% and the 5% positions, and between the 25% and the 50% positions (percentage values indicate location along the duct from the basal end; Fig. 2). These results demonstrate that for OHCs there is limited improvement in orientation over developmental time. In contrast, deviations for IHCs occur in a gradient that extends from base (best orientations) to apex (worst orientations). This gradient parallels the overall basal-to-apical gradient of differentiation within the cochlea and suggests that some improvements in orientation may occur in IHCs. The bases for these differences are unclear, but the observed difference between IHCs and OHCs suggests that *Vangl2* may act differently in each cell type.

Within each hair cell, the first sign of polarization is the movement of the cilium that will give rise to the kinocilium from the centre towards the peripheral edge of the luminal surface. To determine whether *Vangl2* has a role in the initial polarization of developing stereociliary bundles, we analysed changes in the position of the single tubulin-based cilium in each bundle (Fig. 1a, blue dot) in E16.5 cochleae. In the less-mature apical region of wild-type cochleae, cilia on developing OHCs are located near the centre of the cell (Fig. 3a). In contrast, in the more-mature basal region the positions of cilia on OHCs are biased towards the abneural edge (Fig. 3b). In *Lp/Lp* cochleae, the distribution of cilia in the apical region appears centrally located and similar to the wild type (Fig. 3c). In the basal region, although most cilia have moved towards the periphery of the cell, the biased movement towards the abneural edge is disrupted (Fig. 3d). Similar differences are observed for the locations of IHC cilia. In addition, the extent of disruption in the biased movement of OHC cilia occurs in a gradient that parallels the overall neural-to-abneural gradient in bundle deviation (data not shown). These results suggest that the earliest steps in hair cell polarization are disrupted in *Lp/Lp* mutants, and that *Vangl2* regulates the direction of movement and/or anchoring of the cilia (Fig. 3e, f). Moreover, the demon-

stration that initial defects in the direction of cilia movement parallels the degree of mis-orientation in the mature bundle is consistent with the hypothesis that initial defects in OHC bundle orientation are not corrected over time (Fig. 3e, f).

To confirm that *Vangl2* is expressed in the cochlea during stereociliary bundle development, *in situ* hybridization was performed on cochlear sections from E14.5 and E16.5 mice. Messenger RNA for *Vangl2* is present throughout the cochlear epithelium at both stages (Fig. 3g, h). We also observed expression in the spiral ganglion.

As noted, *Lp/Lp* mutants exhibit an open neural tube defect referred to as craniorachischisis. Similar defects are observed in circletail (*Crc*) mutants^{10–12}, which contain a truncation mutation in *Scrb1* (ref. 13), a homologue of *Drosophila scribble*. *Scrb1* is a member of the LAP protein family that is characterized by the presence of four PDZ domains, two of which are absent in *Crc* mutants¹³. To determine whether *Scrb1* also has a role in PCP, stereociliary bundle orientations were determined in *Crc/Crc* cochleae. Defects in stereociliary polarization are present in the ears of *Crc/Crc* mice; however, significant deviations are only observed in second and third row OHCs (Fig. 3i, j). Orientations

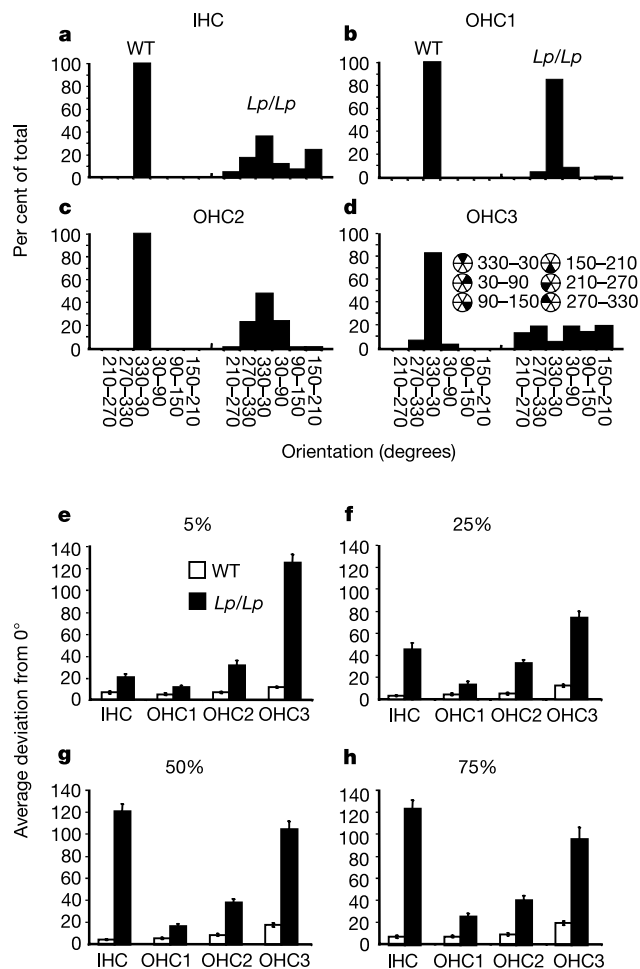


Figure 2 Changes in distribution and average bundle orientation in *Lp/Lp* cochleae. **a–d**, Distribution histograms for orientations in wild-type (WT) and *Lp/Lp* cochleae. In wild-type cochleae, orientations are entirely (IHCs and first and second row OHCs) or largely (third OHCs) confined to a 60° segment centred on a line parallel to the neural–abneural axis. In contrast, the distributions of orientations are noticeably broader in *Lp/Lp* mice. **e–h**, Average deviations from a line parallel to the neural–abneural axis for bundles located at specific positions along the basal–apical axis of the cochlea. Deviations are significant ($P < 0.001$) for all values except OHC1 at 5%.

for IHCs, which are significantly mis-oriented in *Lp/Lp* cochleae, are normal in *Crc/Crc* cochleae (Fig. 3j). To determine whether *Vangl2* and *Scrb1* interact genetically in the regulation of PCP, bundle orientations were analysed in *Lp/+ Crc/+* double heterozygotes. Orientation defects in *Lp/+ Crc/+* mice were comparable with *Lp/Lp* mice, and included defects in both IHCs and OHCs (Fig. 3k, l). Cochleae from animals that were heterozygous for either *Lp* or *Crc* were essentially normal (data not shown). These observations and the co-localization of *Vangl2* (Fig. 3g, h) and *Scrb1* (ref. 13) in the cochlea indicate that these proteins interact to regulate PCP in mammals.

PDZ domain proteins are best known for their roles in sub-membranous receptor and signal transduction complex clustering¹⁴. The presence of four PDZ domains in *Scrb1* indicates a scaffolding function, which might mediate the formation of a multiprotein complex, resulting in the targeting of specific proteins to the apical domain of the hair cell to initiate and/or control PCP.

Recent results have suggested that, in vertebrates, the PCP

pathway also regulates cellular convergent extension^{15,16}. Furthermore, cellular convergent extension has been suggested to have a role in the outgrowth and morphogenesis of the mammalian cochlea¹⁷. Consistent with these data, the bony labyrinth in *Lp/Lp* mice appears shorter and collapsed as compared with wild-type mice (Fig. 4c). In addition, the average length of the cochlear duct is $3,649 \pm 46 \mu\text{m}$ in E18.5 *Lp/Lp* mice ($n = 4$), 27% shorter than in wild-type littermates ($5,020 \pm 125 \mu\text{m}$; $n = 4$) ($P < 0.001$). By comparison, the average number of hair cells in E18.5 *Lp/Lp* cochleae is $2,254 \pm 4$ ($n = 5$), 11% fewer than in wild-type mice ($2,511 \pm 16$; $n = 4$) ($P < 0.001$). As an apparent consequence of the shortening of the cochlear duct there is an atypical accumulation of hair cells in the apex and a subsequent disruption of cellular patterning in the sensory epithelium (see Fig. 4d, e). The basis for the decrease in hair cell number is not clear but does not seem to be a result of increased cell death, as labelling of apoptotic cells indicates no difference between *Lp/Lp* and wild-type cochleae (data not shown).

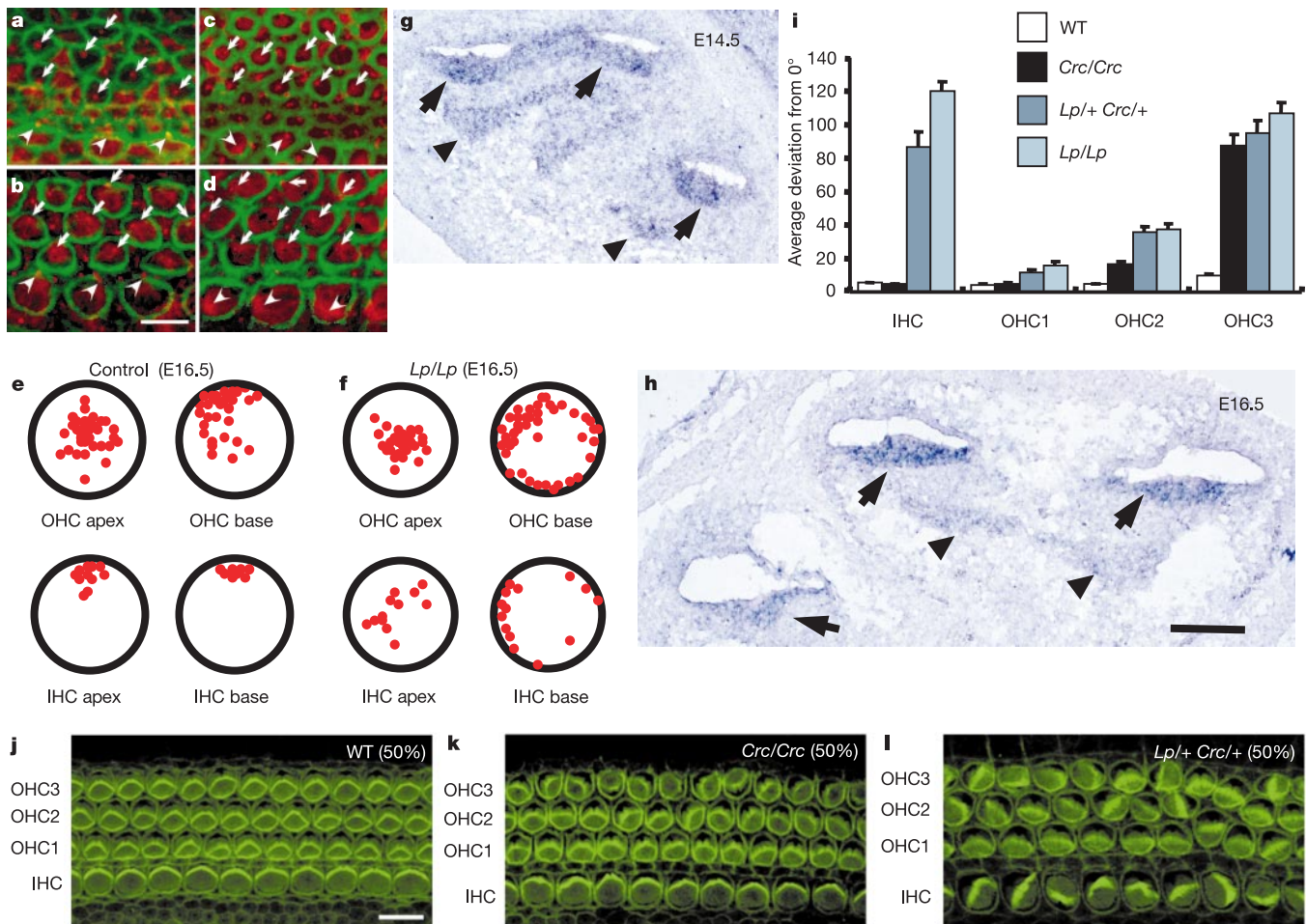


Figure 3 *Vangl2* regulates localization of the kinocilium. **a**, Luminal surface of the apical turn from an E16.5 wild-type cochlea. The developing kinocilia (red) on many OHCs are located near the centre of the cell (arrows). In contrast, the kinocilia on most IHCs are located near the abneural edge (arrowheads). **b**, Same as **a** but for the basal turn. Kinocilia on IHCs and OHCs have moved towards the abneural edge. **c**, Luminal surface of the apical turn from an E16.5 *Lp/Lp* cochlea. The distributions of kinocilia on both OHCs and IHCs are similar to wild-type cochleae. **d**, Same as **c** but for the basal turn. Kinocilia have moved towards the periphery, but the net direction of movement towards the abneural edge is disrupted. **e**, Summary of positions for kinocilia in IHCs and OHCs in E16.5 wild-type cochleae. Each red dot indicates the position of the kinocilium for a single cell. Net movement of kinocilia towards the abneural edge occurs in both IHCs and OHCs.

f, Same as **e** but in *Lp/Lp* cochleae. Movement of kinocilia towards the periphery is not disrupted; however, the direction of movement is no longer biased towards the abneural edge. **g, h**, mRNA for *Vangl2* is expressed in the floor of the cochlear duct (arrows) and in the spiral ganglion (arrowheads) at E14.5 (**g**) and E16.5 (**h**). **i-l**, The organ of Corti in the middle turns of E18 cochleae from wild-type (**j**), *Crc/Crc* (**k**) and *Lp/+ Crc/+* mice (**l**). In the *Crc/Crc* cochlea, orientations on IHCs and OHC1 appear normal, however a clear disruption is present in OHC2 and OHC3. In contrast, marked disruptions in orientations are present in all rows in the *Lp/+ Crc/+* cochlea. **i**, Average deviations for bundles in *Crc/Crc* and *Lp/+ Crc/+* cochleae (*Lp/Lp* data from Fig. 3 are included). Deviations were significant ($P < 0.001$) for IHC and OHC1 values for *Lp/+ Crc/+* but not for *Crc/Crc* mice. Phalloidin label is green; myosin VIIa label is red. Scale bar: **a-d, j-l**, 10 μm ; **e, f**, 20 μm .

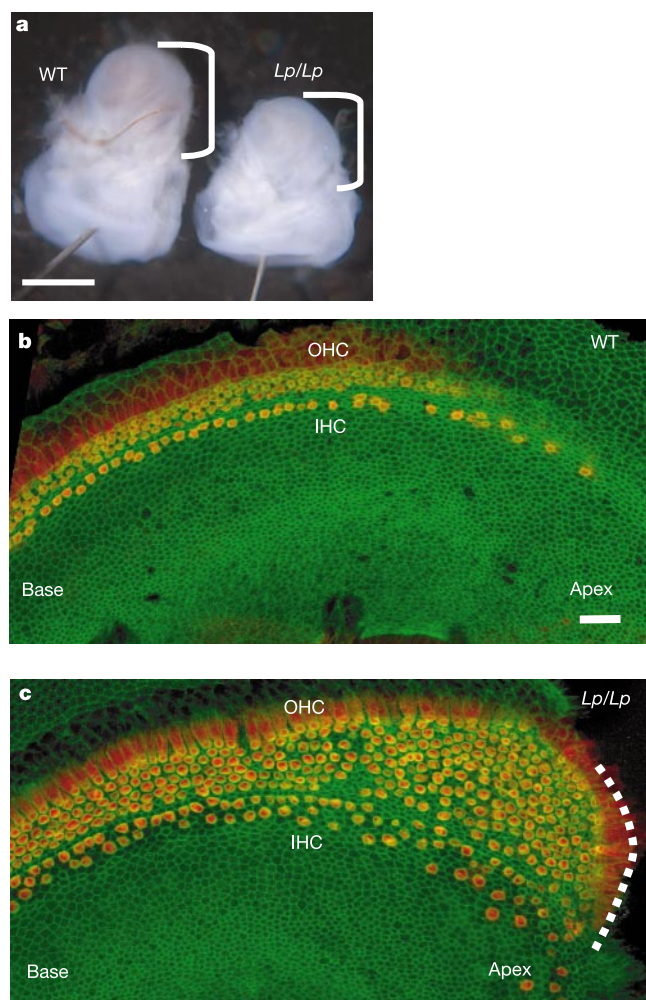


Figure 4 Temporal bones and cochleae are shorter in *Lp/Lp* mice. **a**, Temporal bones from E18.5 wild-type and *Lp/Lp* mice. The cochlear portion (brackets) is smaller in *Lp/Lp* mice. **b**, Apical region of the organ of Corti in an E18.5 wild-type cochlea. The number of hair cells is significantly reduced near the end of the duct. **c**, In contrast, in an *Lp/Lp* cochlea there is an accumulation of both IHCs and OHCs that extends to the end of the duct (dotted line). Scale bar: **a**, 1 μ m; **b**, **c**, 20 μ m.

Our results provide evidence for a conservation of function for the PCP pathway in the generation of planar polarity between *Drosophila* and mammals. Of note, mice deficient for *dvl1* and *dvl2* (homologues of *Drosophila dsh*) develop with a neural tube phenotype similar to *Lp/Lp* and *Crc/Crc* homozygotes¹⁸. As *dsh* is an important upstream component of the PCP pathway in *Drosophila*, this suggests the existence of a common PCP-like pathway in mammals. In contrast with *Stbm/Vang*, *Scribble* does not have a role in *Drosophila* PCP; rather, it regulates the apical-basal polarity of epithelial cells by dictating the proper confinement of apical proteins¹⁹. Therefore, it seems possible that in mammals the correct localization and function of *Scrb1* is a prerequisite for the proper localization—and consequently function—of PCP proteins, such as *Vangl2*. These results also demonstrate that although the core PCP genes seem to be conserved between vertebrates and invertebrates, important differences exist. Finally, our results also support the hypothesis that cellular convergent extension has a role in the development and patterning of the cochlear duct, and suggest that the role of *Stbm/Vang* in cellular convergent extension has been conserved among vertebrate classes^{20,21}.

The mechanism(s) by which *Vangl2* and *Scrb1* regulate bundle orientation are unclear. Our data and that of others¹³ suggest that

these proteins could associate, probably with other proteins, as a complex to influence the position and/or anchoring of the kinocilium and the subsequent orientation of the bundle. The observation that bundle orientations are less affected in the first and second row OHCs, as compared with third row OHCs and IHCs, suggests that these cells are less dependent on *Vangl2* and *Scrb1* for the determination of bundle orientation. These results suggest that unique and complex mechanisms are involved in the determination of orientation in IHCs compared with OHCs. □

Methods

Loop-tail and circletail mice

Loop-tail mutant mice of the LPT/Le stock were obtained from Jackson Laboratories. The circletail mutation arose spontaneously in 1997 in a transgenic colony at Columbia University¹². Both strains of mice were maintained at NIH under standard conditions by heterozygous intercross. To generate homozygous mutants, heterozygous animals were crossed, and timed pregnant litters were obtained at E16.5, E18 and E18.5. Heterozygous offspring were identified by a looped tail, whereas homozygous fetuses die perinatally and were distinguished embryonically by the presence of craniorachischisis^{7,22}.

Immunostaining, SEM and histology

Temporal bones were dissected from mutant embryos at the stated ages. After removal of the surrounding cartilage, the anlage of Reissner's membrane was dissected to expose the developing sensory epithelium. To determine the number of hair cells and to quantify stereociliary bundle deviations, cochleae from wild-type, *Lp/Lp*, *Crc/Crc* and *Lp/Lp* + *Crc/Crc* mice were labelled with the hair cell markers myosin VI (gift from T. Hasson) or myosin VIIa (gift from C. Petit), and filamentous actin was labelled with phalloidin (Alexa Fluor 488-conjugated; Molecular Probes) as reported previously²³. Individual cilia were labelled with anti- β -tubulin-rhodamine antibody (Sigma). For SEM, tissues were processed as reported previously²⁴. To examine cellular histology, E18.5 wild-type and *Lp/Lp* cochleae were fixed, embedded, sectioned and stained as described previously²⁴.

In situ hybridization

Pregnant CD-1 mice (Charles River Laboratories) were killed according to the NIH Guide for the Care and Use of Laboratory Animals. Embryos were staged individually, and cochleae were dissected, cryosectioned at a thickness of 12 μ m, and processed for *in situ* hybridization as previously described²⁵. A digoxigenin-labelled antisense RNA probe for *Vangl2* was generated from a pCMV SPORT6 vector carrying a 1-kilobase fragment of the coding region (expressed sequence tag BF118609).

Analysis of stereociliary bundle orientation

Cochleae from E16.5, E18 and E18.5 mice were labelled as described and then flat mounted. The entire length of the cochlear duct was determined and positions located 5%, 25%, 50% and 75% from the basal end of the duct were identified. For each position, an image of the organ of Corti containing approximately 20 IHCs and 60 OHCs was obtained. Orientations for individual stereociliary bundles were determined relative to a line parallel to the neural-abneural axis and perpendicular to the row of pillar cells (see Fig. 1). The vertex of the symmetric bundle was used to determine the orientation for each bundle, and cells in which the vertex was exactly parallel with the neural-abneural axis were assigned an orientation of 0°. Distribution histograms were generated by determining the orientations for bundles ($n > 50$ for each hair cell type) based on a clockwise rotation towards the apex of the cochlea. For average deviations, the absolute rotation away from the 0° orientation was determined regardless of direction.

Received 13 February; accepted 28 March 2003; doi:10.1038/nature01618.

Published online 30 April 2003.

- Adler, P. N. Planar signaling and morphogenesis in *Drosophila*. *Dev. Cell* **2**, 525–535 (2002).
- Adler, P. N. & Lee, H. Frizzled signaling and cell-cell interactions in planar polarity. *Curr. Opin. Cell Biol.* **13**, 635–640 (2001).
- Taylor, J., Abramova, N., Charlton, J. & Adler, P. N. *Van Gogh*: a new *Drosophila* tissue polarity gene. *Genetics* **150**, 199–210 (1998).
- Wolff, T. & Rubin, G. M. *Strabismus*, a novel gene that regulates tissue polarity and cell fate decisions in *Drosophila*. *Development* **125**, 1149–1159 (1998).
- Murdoch, J. N., Doudney, K., Paternotte, C., Copp, A. J. & Stanier, P. Severe neural tube defects in the loop-tail mouse result from mutation of *Lpp1*, a novel gene involved in floor plate specification. *Hum. Mol. Genet.* **10**, 2593–2601 (2001).
- Kibar, Z. *et al.* *Ltap/Lpp1*, a mammalian homolog of *Drosophila* Strabismus/Van Gogh, is altered in the mouse neural tube mutant Loop-tail. *Nature Genet.* **28**, 251–255 (2001).
- Strong, L. C. & Hollander, W. F. Hereditary loop-tail in the house mouse accompanied by inperforate vagina and with lethal craniorachischisis when homozygous. *J. Hered.* **40**, 329–334 (1949).
- Hudspeth, A. J. & Jacobs, R. Stereocilia mediate transduction in vertebrate hair cells (auditory system/cilium/vestibular system). *Proc. Natl Acad. Sci. USA* **76**, 1506–1509 (1979).
- Rubel, E. W. in *Handbook of Sensory Physiology IX* (ed. Jacobson, M.) 135–237 (Springer, New York, 1978).
- Rachel, R. A., Murdoch, J. N., Beermann, F., Copp, A. J. & Mason, C. A. Retinal axon misrouting at the optic chiasm in mice with neural tube closure defects. *Genesis* **27**, 32–47 (2000).
- Murdoch, J. N. *et al.* Circletail, a new mouse mutant with severe neural tube defects: chromosomal localization and interaction with the loop-tail mutation. *Genomics* **78**, 55–63 (2001).
- Rachel, R. A., Wellington, S. J., Warburton, D., Mason, C. A. & Beermann, F. A new allele of Gli3 and a new mutation, circletail (*Crc*), resulting from a single transgenic experiment. *Genesis* **33**, 55–61 (2002).

13. Murdoch, J. N. *et al.* Disruption of scribble (Scrb1) causes severe neural tube defects in the circletail mouse. *Hum. Mol. Genet.* **12**, 87–98 (2003).
14. Hung, N. Y. & Sheng, M. PDZ domains: structural molecules for protein complex assembly. *J. Biol. Chem.* **277**, 5699–5702 (2002).
15. Mlodzik, M. Planar cell polarization: do the same mechanisms regulate *Drosophila* tissue polarity and vertebrate gastrulation? *Trends Genet.* **18**, 564–571 (2002).
16. Park, M. & Moon, R. T. The planar cell-polarity gene *stbm* regulates cell behaviour and cell fate in vertebrate embryos. *Nature Cell Biol.* **4**, 20–25 (2002).
17. Chen, P., Johnson, J. E., Zoghbi, H. Y. & Segil, N. The role of Math1 in inner ear development: uncoupling the establishment of the sensory primordium from hair cell fate determination. *Development* **129**, 2495–2505 (2002).
18. Hamblet, N. S. *et al.* Dishevelled 2 is essential for cardiac outflow tract development, somite segmentation and neural tube closure. *Development* **129**, 5827–5838 (2002).
19. Bilder, D. & Perrimon, N. Localization of apical epithelial determinants by the baso-lateral PDZ protein scribble. *Nature* **403**, 676–680 (2000).
20. Darken, R. S. *et al.* The planar polarity gene *strabismus* regulates convergent extension movements in *Xenopus*. *EMBO J.* **21**, 976–985 (2002).
21. Goto, T. & Keller, R. The planar cell polarity gene *strabismus* regulates convergence and extension and neural fold closure in *Xenopus*. *Dev. Biol.* **247**, 165–181 (2002).
22. Copp, A. J., Checiu, I. & Henson, J. N. Developmental basis of severe neural tube defects in the *loop-tail* (*Lp*) mutant mouse: use of microsatellite DNA markers to identify embryonic genotype. *Dev. Biol.* **165**, 20–29 (1994).
23. Mueller, K. L., Jacques, B. E. & Kelley, M. W. Fibroblast growth factor signaling regulates pillar cell development in the organ of Corti. *J. Neurosci.* **22**, 9368–9377 (2002).
24. Lanford, P. J. *et al.* Notch signalling pathway mediates hair cell development in mammalian cochlea. *Nature Genet.* **21**, 289–292 (1999).
25. Wu, D. K. & Oh, S. H. Sensory organ generation in the chick inner ear. *J. Neurosci.* **16**, 6454–6462 (1996).

Acknowledgements We thank J. Jones for assistance with *in situ* hybridization and C. Woods for technical support. We also thank R. Wenthold, N. Sans, I. Belyantseva, D. Forrest, A. Dabdoub, J. Jones and R. Hertzano for reading an earlier version of this manuscript. This research was supported by funds from the Intramural Programs at NIDCD and NCI.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.M. (montcouq@nidcd.nih.gov).

Pathogenic bacteria attach to human fibronectin through a tandem β -zipper

Ulrich Schwarz-Linek*, Jörn M. Werner*, Andrew R. Pickford*, Sivashankarappa Gurusiddappa†, Jung Hwa Kim†, Ewa S. Pilka*, John A. G. Briggs‡, T. Sebastian Gough*, Magnus Höök†, Iain D. Campbell*§ & Jennifer R. Potts*

* Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK

† Center for Extracellular Matrix Biology, Institute of Biosciences and Technology, Texas A&M University System Health Science Center, Houston, Texas 77030-3303, USA

§ Oxford Centre for Molecular Sciences, Central Chemistry Laboratory, University of Oxford, South Parks Road, Oxford OX1 3QH, UK

Staphylococcus aureus and **Streptococcus pyogenes**, two important human pathogens, target host fibronectin (Fn) in their adhesion to and invasion of host cells^{1,2}. Fibronectin-binding proteins (FnBPs), anchored in the bacterial cell wall, have multiple Fn-binding repeats³ in an unfolded^{4,5} region of the protein. The bacterium-binding site in the amino-terminal domain (^{1–5}F1) of Fn contains five sequential Fn type 1 (F1) modules. Here we show the structure of a streptococcal (*S. dysgalactiae*) FnBP peptide (B3)^{6,7} in complex with the module pair ¹F1²F1. This identifies ¹F1- and ²F1-binding motifs

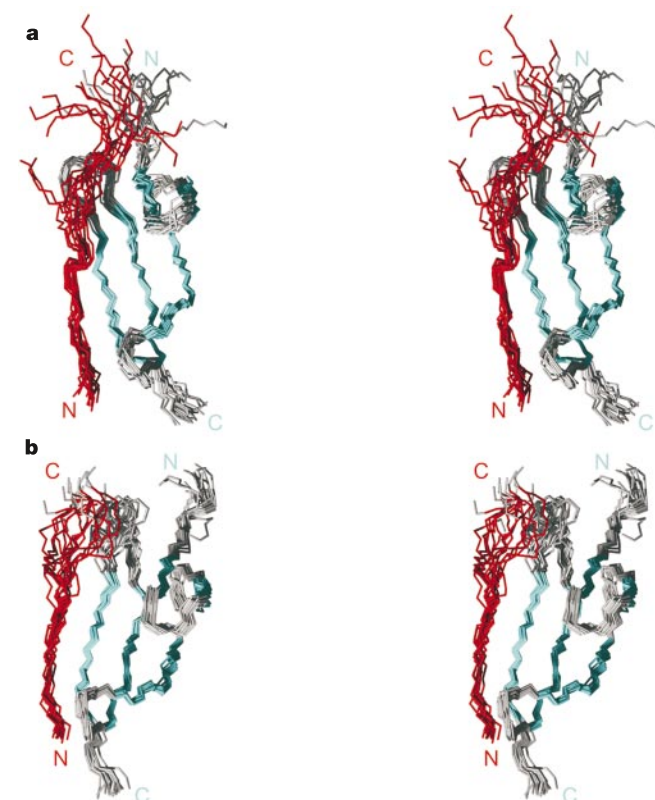


Figure 1 Family of NMR-derived structures of the ¹F1²F1 module pair from human Fn complexed with B3 from *S. dysgalactiae* (Table 1). Stereo view of structures overlaid on the backbone C α atoms of the F1 β -strands of ¹F1–B3(1052–1066) (a) and ²F1–B3(1043–1051) (b). F1 β -strands and B3 residues are shown in cyan and red, respectively. The N and C termini of ¹F1, ²F1 (cyan) and B3 regions (red) are indicated.

in B3 that form additional antiparallel β -strands on sequential F1 modules—the first example of a tandem β -zipper. Sequence analyses of larger regions of FnBPs from *S. pyogenes* and *S. aureus* reveal a repeating pattern of F1-binding motifs that match the pattern of F1 modules in ^{1–5}F1 of Fn. In the process of Fn-mediated invasion of host cells, therefore, the bacterial proteins seem to exploit the modular structure of Fn by forming extended tandem β -zippers. This work is a vital step forward in explaining the full mechanism of the integrin-dependent^{2,8} FnBP-mediated invasion of host cells.

Fn and Fn–integrin interactions are required for normal physiological processes. Bacterial FnBP-mediated adhesion to and integrin-dependent invasion of host cells therefore seem to be examples of the exploitation of host cell processes by pathogens⁹. Internalization might aid evasion of the immune system and administered antibiotics by allowing bacteria to ‘hide’ inside host cells. In addition, as bacteria must cross the endothelial lining to enter or leave the vasculature, internalization might assist the haematogenous dissemination of infection to other tissues, which is a characteristic of invasive infection by *S. aureus*¹⁰. In support of this, FnBPs are central to the interaction of *S. aureus* with endothelial cells *in vitro*^{1,8}. FnBPs also mediate the attachment of bacteria to implanted material, which becomes coated with host proteins¹¹, indicating a possible role in infections associated with catheters and prosthetic devices. The absence of three-dimensional structural information until now has precluded an explanation of the mechanism by which the largely unstructured Fn-binding repeats of FnBPs^{4,5} recognize ^{1–5}F1.

The structure of the ¹F1²F1–B3 complex (Fig. 1) was determined by NMR spectroscopy. In the family of structures (Fig. 1), the

‡ Present address: Division of Structural Biology, The Wellcome Trust Centre for Human Genetics, Roosevelt Drive, Headington, Oxford OX3 7BN, UK.

individual F1 modules are well defined (Table 1) and retain the structure of the free F1 module, characterized by a β -sandwich of two antiparallel β -sheets; a double-stranded sheet (strands A and B) followed by a triple-stranded sheet (strands C, D and E; Fig. 2)¹². On binding to the module pair, the bacterial peptide contributes a fourth antiparallel strand to the triple-stranded β -sheet of sequential F1 modules in a tandem β -zipper interaction. In NMR spectra of the complex, $H\alpha$ - $H\alpha$ intermolecular distances typical of β -sheet formation were observed for residues shown in italics in the sequences below.

The complex reveals two different F1 recognition sequences within B3 and their molecular character is mirrored by the properties of the two F1 interaction surfaces. The carboxy-terminal LSIHFDNEWP part of the B3 peptide binds along a wide groove on the surface of the ¹F1 module bounded by the C-D and D-E loops (Fig. 2a). The alternating hydrophobic side chains of the peptide residues underlined above interact with a hydrophobic patch formed largely by Tyr 38 in the C-D loop and Leu 43 in strand D. Additional interactions are observed between Tyr 22 in the A strand of ¹F1 and the B3 residues in bold above. The presence of the [I/F]HFDNX(X)P motif in streptococcal FnBPs has been noted previously¹³. However, as this sequence only occurs once within each protein it has not usually been considered part of the F1-binding repeat regions. The ¹F1²F1-B3 structure reveals the role of this motif in binding ¹F1.

In comparison with ¹F1, the surface of the peptide-binding region of ²F1 is more basic (Fig. 2a). This is due to the presence of Lys 85 in the C-D loop of ²F1 (in a homologous position to Tyr 38 in ¹F1), Arg 101 in the D-E loop and Lys 69 in the A strand¹². The region of B3 binding to ²F1 contains the sequence STTEVED[S/T]. The residues in bold are conserved in FnBPs from Gram-positive bacteria⁶ and in the structure are in close proximity to the positively charged region of the ²F1 surface (Fig. 2a); thus electrostatic interactions are likely to be important in the binding. The residues underlined above in the ¹F1- and ²F1-binding regions of B3 are highlighted in Fig. 2a.

The directionality of the peptide binding is encoded in the antiparallel β -sheet extension of the individual F1 modules and can be best appreciated in the structure of the module pair. Figure 2 shows an elongated conformation for the ¹F1²F1-B3 complex as is observed in most (14 of 15) structures in the family shown in Fig. 1. The inter-module orientation is significantly better defined in the complex with B3 than in the free form¹², indicating that B3 might

tether the two F1 modules. However, a degree of variability remains in the inter-module orientation even after the inclusion in the structure calculations of long-range restraints based on residual dipolar couplings and ¹⁵N-relaxation rates.

The structure reveals (1) β -strand formation (a β -zipper) as a mechanism of F1 module recognition; (2) an antiparallel orientation for Fn and FnBP; (3) the approximate length (about eight residues) of an F1-binding motif; (4) likely sources of F1-binding specificity in the F1 and peptide sequences; and (5) that B3 forms a tandem β -zipper on adjacent F1 modules. The combination of these results with sequence analyses of homologous FnBPs and previous results^{7,14} leads to a novel model for binding of *S. pyogenes* and *S. aureus* to ¹⁻⁵F1 (Fig. 3).

The C-terminal regions of the FnBPs SfbI (from *S. pyogenes*)¹⁵ and FnBPA (from *S. aureus*)¹⁶ can be arranged into 5 and 11 segments (SfbI-1 to SfbI-5 and FnBPA-1 to FnBPA-11), respectively, such that each (except FnBPA-7) contains a similar series of putative F1-binding motifs. The definition of the F1-binding motifs is assisted by sequence conservation between the segments and between SfbI, FnBPA and other FnBPs (data not shown), but it is clear that the sequences of some motifs are more conserved than others. Each segment in SfbI contains motifs in the correct order to bind sequential F1 modules in ²⁻⁵F1 or ¹⁻⁵F1. The motifs are predicted to bind F1 modules primarily by forming a β -strand along the E strand of the triple-stranded β -sheet in a similar manner to the B3 interaction with ¹F1²F1 shown in Fig. 1. However, in the

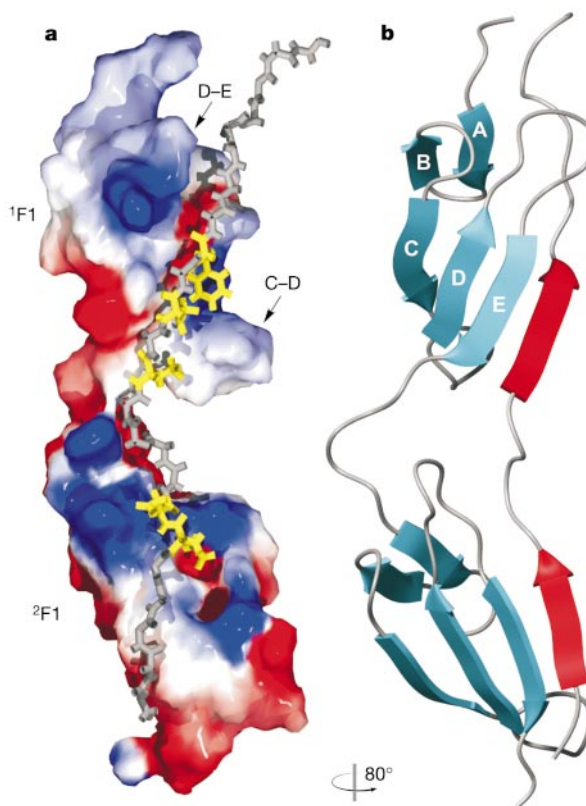


Figure 2 Molecular surface (prepared using GRASP) and ribbon diagram (prepared with MOLMOL) of ¹F1²F1-B3. **a**, Surface potential of ¹F1²F1 shown on the lowest-energy structure from Fig. 1 with bound B3 (grey). Negatively and positively charged regions of the ¹F1²F1 surface are shown in red and blue, respectively. Side chains of hydrophobic and acidic B3 residues (yellow) are mentioned in the text. **b**, Ribbon diagram of the lowest-energy structure showing β -strands of the F1 modules (cyan) and the fourth strand formed by B3 (red). The definition of the sheet extension is based on characteristic cross-peaks in NOESY spectra and changes in ¹H NMR chemical shifts²⁹. The difference in orientation between the two views is indicated. F1 β -strand labels are shown on ¹F1.

Table 1 Experimental restraints and structural statistics

Number of experimental restraints	¹ F1 ² F1	B3	Intermolecular
NOEs	1279	157	127
Hydrogen bonds	7		
TALOS (ϕ/ψ)	76		
¹⁵ N relaxation restraints	29		
¹⁵ N- ¹ H RDC restraints	40		
R.m.s.d. from experimental restraints (means $\pm$ s.d.)			
Distance ($\text{\AA}$)	0.020 $\pm$ 0.002		
Angles (degrees)	0.39 $\pm$ 0.08		
Hydrogen bond ($\text{\AA}$)	0.010 $\pm$ 0.006		
R.m.s.d. from idealized covalent geometry (means $\pm$ s.d.)			
Bonds ($\text{\AA}$)	0.0020 $\pm$ 0.0001		
Angles (degrees)	0.319 $\pm$ 0.016		
Impropers (degrees)	0.38 $\pm$ 0.02		
R.m.s.d. of cartesian coordinates (means $\pm$ s.d.)			
	¹ F1-B3	² F1-B3	¹ F1 ² F1*
Backbone ($\text{\AA}$)	0.50 $\pm$ 0.08	0.49 $\pm$ 0.12	1.13 $\pm$ 0.31
All heavy atoms ($\text{\AA}$)	0.97 $\pm$ 0.10	0.96 $\pm$ 0.20	1.42 $\pm$ 0.29

Fifteen structures were chosen on the basis of low overall energy. None of the structures showed experimental distance or angle violations of more than 0.3 $\text{\AA}$ or more than 3°, respectively. Root-mean-square deviation (r.m.s.d.) calculations were performed over β -strand residues of the F1 modules¹² and for B3 residues 23-28 (binding to ¹F1) and 13-18 (binding to ²F1). The 14 low-energy structures showed an elongated conformation in the complex.

*Structure 15 had a different intermodule orientation, resulting in lower consistency with RDC restraints. This structure was not included in the overlay of both modules.

¹F1- and ²F1-binding motifs the β -strand-forming residues are not the most conserved, indicating that additional F1-binding specificity might arise from interactions of the side chains of conserved residues with the F1 modules. This is supported by the observation of interactions between B3 and ¹F1 A-strand residues (Fig. 1) in the NMR spectra of the ¹F1²F1-B3 complex¹⁷ (see above). In FnBPA, the pattern of motifs and previous data^{6,14} suggest a similar mechanism of binding to the N-terminal domain of Fn. However, there is no obvious ¹F1-binding motif in FnBPA, and the lack of sequence conservation indicates that some segments might not bind ⁵F1.

The extended tandem β -zipper model described above was tested experimentally (Fig. 3). Peptides from SfbI containing putative ¹F1²F1-, ²F1³F1- and ⁴F1⁵F1-binding sites were shown to bind to their predicted module pair targets. When the predicted binding sites *in vivo* are intact (SfbI-5 and ¹⁻⁵F1), the lower-affinity interactions involving F1 module pairs combine to form a high-affinity multipoint interaction (Fig. 3). It should be noted that the interaction between SfbI-4 or SfbI-5 and ¹⁻⁵F1 is weaker than predicted by the product of the K_d values (measured at low ionic strength) for the interactions of the subdomains. This is likely to be due to the dependence of the interactions on ionic strength (data not shown). At lower ionic strength (see Methods) the affinity of the interaction between SfbI-4 or SfbI-5 and ¹⁻⁵F1 would have been too high to measure directly by isothermal titration calorimetry (ITC). Syn-

thetic peptides from FnBPA were also shown to bind to ²F1³F1 and ⁴F1⁵F1 as predicted. The specificity of the ⁴F1⁵F1-binding peptide has been demonstrated previously^{7,18}. The ²F1³F1-binding motif from FnBPA-1 is outside the traditionally defined Fn-binding repeats in FnBPA¹⁶ (Fig. 3) and is a strong test of the model. The existence of binding sites for Fn near the C terminus of the A domain (Fig. 3) has been demonstrated recently¹⁹.

In addition to the ¹F1²F1-B3 structure, further evidence of the formation of a β -zipper along the E strand of the F1 modules comes from changes in ²F1³F1 (data not shown) and ⁴F1⁵F1 (ref. 14) backbone NMR chemical shifts when the FnBP peptide binds. In the model (Fig. 3), SfbI and FnBPA contain multiple binding sites for ²⁻⁵F1. SfbI also contains a single ¹⁻⁵F1-binding site near the C terminus of the protein. Thus, the extended tandem β -zipper model is consistent with earlier studies indicating that D1-D3 from FnBPA (Fig. 3) binds two molecules of ¹⁻⁵F1 with high (nanomolar) affinity¹⁸ and that FnBPA contains multiple substituting binding sites for Fn¹⁹. The common themes in streptococcal and staphylococcal Fn binding identified here (despite the lack of an obvious ¹F1-binding motif in FnBPA) are consistent with the ability of peptides from streptococcal FnBPs to inhibit the binding of *S. aureus* to Fn⁶. The model also provides a plausible mechanism for the propagation of a conformational change along SfbI that could explain the apparent activation of UR domain (Fig. 3) binding to the Fn gelatin-binding domain²⁰.

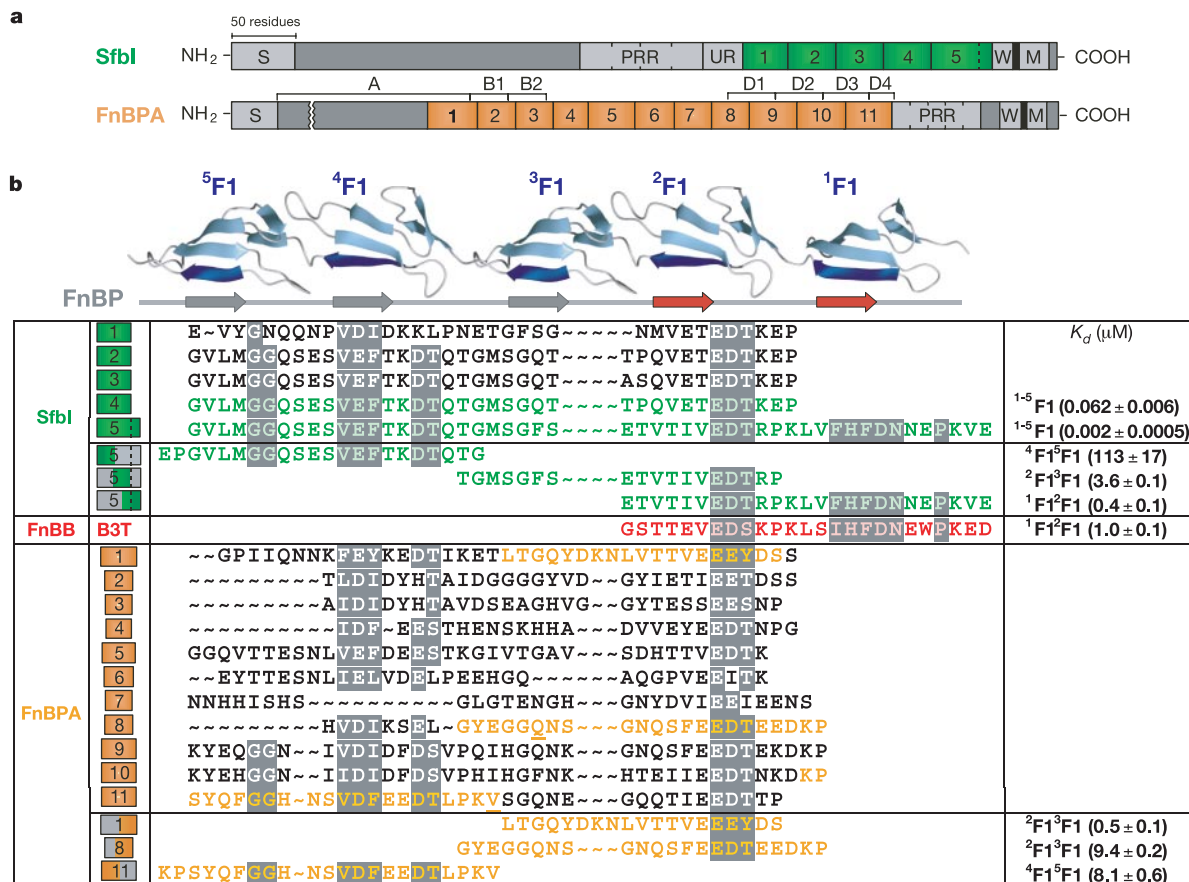


Figure 3 The extended tandem β -zipper model. **a**, Schematic representations of SfbI^{15,30} and FnBPA¹⁶ showing putative ¹⁻⁵F1-binding segments (green and orange boxes). For FnBPA, the locations of the traditionally defined regions (A, B and D) are also shown. The first and last residues of D1-D3 are underlined in **b**. Other terms are explained in Talay *et al.*³⁰ and Swiss-Prot entry P14738. **b**, The extended tandem β -zipper model is at the top. A ¹⁻⁵F1-binding segment is shown as a grey line with short β -strands along the E strand (shown in navy blue) of five sequential modules with the consensus F1 fold

(ref. 12; cyan). FnBP β -strands for which three-dimensional structural data are available (Fig. 1) are shown in red. The position of the F1 modules approximates the location of F1-specific motifs in the segments aligned below. Peptides used in testing the model and the K_d values for the interactions are shown. The location of these peptides in the segments is also indicated. B3T (red) contains all the ¹F1²F1-binding residues of B3 and binds with similar affinity¹⁷.

The formation of intermolecular or subunit interfaces through the extension of a β -sheet (a β -zipper interaction) has been observed on several occasions, for example in an interaction involving an IgG-binding protein²¹. Here we have identified the first tandem β -zipper interaction, where β -strands are formed on adjacent domains, and we also provide evidence that this is part of an extended tandem β -zipper that forms on the binding of larger regions of the bacterial and host proteins. As well as showing how pathogenic bacteria can adhere to F1 modules in $^{1-5}$ F1 of Fn, the consequences of the mechanism, namely the high affinity and specificity of the binding, the significant conformational change undergone by the unfolded region of the FnBP and the presence of multiple Fn-binding sites, are all likely to have a role in bacterial invasion. This work is therefore a significant step towards the goal of understanding this process. $\square$

Methods

Recombinant expression and peptide synthesis

Production of the *P. pastoris* clones expressing $^1\text{F1}^2\text{F1}$ and $^4\text{F1}^5\text{F1}$ (residues 17–109 and 152–244, respectively, of mature human Fn) have been described previously¹². The clone expressing $^2\text{F1}^3\text{F1}$ (residues 62–151 of mature human Fn) was produced in an analogous fashion. Unlabelled $^1\text{F1}^2\text{F1}$, $^2\text{F1}^3\text{F1}$ and $^4\text{F1}^5\text{F1}$, uniformly ^{15}N -labelled $^1\text{F1}^2\text{F1}$ ($[\text{U}-^{15}\text{N}]^1\text{F1}^2\text{F1}$) and $^4\text{F1}^5\text{F1}$ ($[\text{U}-^{15}\text{N}]^4\text{F1}^5\text{F1}$), and uniformly $^{13}\text{C}/^{15}\text{N}$ -labelled $^1\text{F1}^2\text{F1}$ ($[\text{U}-^{13}\text{C}/^{15}\text{N}]^1\text{F1}^2\text{F1}$) were expressed in fermenters by using established methods^{12,22}. For the expression of uniformly $^1\text{H}/^{15}\text{N}$ -labelled $^1\text{F1}^2\text{F1}$ ($[\text{U}-^1\text{H}/^{15}\text{N}]^1\text{F1}^2\text{F1}$), the *P. pastoris* cells were conditioned for growth in perdeuterated medium by sequentially subculturing mid-exponential phase cells from unlabelled buffered minimal glycerol medium²² into medium containing 50%, 80% or 100% $^2\text{H}_2\text{O}$. These cells were fermented in perdeuterated basal salts medium containing *Pichia* trace minerals, 1% (w/v) ($^{15}\text{N}_2\text{H}_4$) $_2\text{SO}_4$ and 0.5% (w/v) ^2H glycerol. On depletion of the carbon source, $^1\text{F1}^2\text{F1}$ expression was induced with 1% (v/v) ^2H methanol. $^1\text{F1}^2\text{F1}$, $^2\text{F1}^3\text{F1}$ and $^4\text{F1}^5\text{F1}$ were purified and characterized much as described previously¹². Segments encoding SfbI-4 and SfbI-5 were amplified by PCR from *S. pyogenes* M75 DNA (4673) using the oligonucleotides 5'-GCGGATCCCGGAGTATTGATGGGAGGTCAAAGT-3', and for SfbI-4: 5'-GCGCTCGAGTCATGGCTCTTCGTGTCCTCTGCTCTC-3', or for SfbI-5: 5'-GCGCTCGAGTCATTCACCTTTGGGCTCATTATTGTC-3'. The amplified PCR product was digested with *Bam*HI and *Xho*I and ligated into appropriately digested pGEX-5X-1 (Amersham Pharmacia Biotech). Transformed *Escherichia coli* JM101 cells with insert were picked and grown on plates made with Luria–Bertani medium containing ampicillin (100 $\mu\text{g ml}^{-1}$; LB-Amp). Recombinant glutathione S-transferase (GST) fusion protein was expressed and purified in accordance with the manufacturer's instructions with slight modification. Overnight culture of transformant was diluted 1:25 in LB-Amp and incubated at 37 °C with shaking until D_{600} reached 0.7. Expression was induced for 4 h by adding isopropyl β -D-thiogalactopyranoside (Invitrogen) to a final concentration of 0.2 mM. The GST fusion protein was purified from lysates on a glutathione–agarose column (Sigma). GST was cleaved off with Factor Xa (New England Biolabs), and the peptide (residues 505–541 or 542–591 of SfbI for SfbI-4 and SfbI-5, respectively, with three non-native N-terminal residues, Gly-Ile-Pro) was purified by reverse-phase high-performance liquid chromatography (HPLC). B3 (EESLPTEQGQSGSTTEVEDSKPKL SIHFDNEWPKED; residues 1031–1066) from *S. dysgalactiae* FnBB and other peptides were synthesized in house or obtained from Alta Bioscience. Residue numbers for the $^1\text{F1}^2\text{F1}$ -, $^2\text{F1}^3\text{F1}$ - and $^4\text{F1}^5\text{F1}$ -binding peptides from SfbI-5 are 567–591, 560–577 and 540–561, respectively. Residue numbers for $^2\text{F1}^3\text{F1}$ -binding peptides from FnBPA-1 and FnBPA-8 are 531–549 and 741–762, respectively. Residue numbers for the $^4\text{F1}^5\text{F1}$ -binding peptide from FnBPA-11 are 837–858. Peptides were purified by reverse-phase HPLC, and the mass was confirmed by electrospray or matrix-assisted laser desorption/ionization mass spectrometry.

Structural restraints

Spectrometers were as described previously¹². Standard two- and three-dimensional experiments, acquired at pH 5.0 and 25 °C or 10 °C, were used for the assignment and collection of restraints for structure calculations. Spectra of $[\text{U}-^{15}\text{N}/^1\text{H}]^1\text{F1}^2\text{F1}$ with different B3 concentrations greatly assisted the assignment of bound B3. The following spectra of $[\text{U}-^{13}\text{C}/^{15}\text{N}]^1\text{F1}^2\text{F1}$:B3T 1:2 were acquired to obtain C' and $\text{C}\alpha$ chemical shifts: HNCO, CCONH, $\{^1\text{H}\}$ constant-time heteronuclear single-quantum coherence (HSQC) and ^{13}C - ^1H nuclear Overhauser enhancement spectroscopy (NOESY)-HSQC. Intramolecular distance restraints for the module pair and peptide were collected from NOESY spectra of samples where $^1\text{F1}^2\text{F1}$ or B3 was fully bound, respectively. Samples in which more than 80% $^1\text{F1}^2\text{F1}$ and more than 98% B3 were bound were used for intermolecular restraints. NOE-based and hydrogen-bond distance restraints were collected and incorporated much as described previously¹². Restraints derived from TALOS (Torsion angle likelihood obtained from shift and sequence similarity)²³ for ϕ and ψ angles were also used. Structure calculations were performed with a version of XPLOR 3.8 (ref. 24) provided by M. Clore. Residual dipolar couplings (RDCs) between backbone ^1H and ^{15}N atoms were measured for the $[\text{U}-^{13}\text{C}/^{15}\text{N}]^1\text{F1}^2\text{F1}$ -B3 complex within the aqueous phase of a polyacrylamide gel (4%) under radial compression²⁵. The H–N couplings were measured as described previously²⁶. Restraints based on ^{15}N relaxation data for the $^1\text{F1}^2\text{F1}$ -B3 complex were measured as described previously²⁷. RDC and ^{15}N

relaxation-based restraints²⁸ were introduced during the refinement stage. One hundred structures were calculated and the 15 lowest-energy structures were selected for analysis. Figures were prepared with MOLMOL (authors R. Koradi, M. Billeter and K. Wüthrich), POV-Ray and MegaPov (authors The POV-Ray team), GRASP (authors A. Nicholls, R. Bharadwaj and B. Honig; with full.crg charge file), Swiss-PdbViewer (authors N. Guex and M. C. Peitsch) and Raster3d (authors E. A. Merritt and M. E. P. Murphy).

Dissociation constants

K_d values for SfbI peptide binding to $^1\text{F1}^2\text{F1}$ (pH 5.0) and $^2\text{F1}^3\text{F1}$ (pH 7.4) and FnBPA peptide binding to $^2\text{F1}^3\text{F1}$ (pH 7.4) and $^4\text{F1}^5\text{F1}$ (pH 5.0) were measured at 25 °C by ITC in low-ionic-strength buffer as described previously¹⁷. The K_d for SfbI-4 or SfbI-5 binding to $^{1-5}\text{F1}$ was measured at physiological ionic strength at 25 °C or 37 °C, respectively, in an ITC experiment with the 30-kDa N-terminal fragment of Fn (7 μM or 2 μM , respectively; Sigma) and SfbI-4 (91 μM) or SfbI-5 (29 μM). The K_d for SfbI peptide binding to $[\text{U}-^{15}\text{N}]^4\text{F1}^5\text{F1}$ was measured by fitting changes in backbone ^1H and ^{15}N NMR chemical shifts to the following equation:

$$\Delta = \Delta_0 \frac{K_d + [\text{L}] + [\text{P}] - \sqrt{(K_d + [\text{L}] + [\text{P}])^2 - 4[\text{L}][\text{P}]}}{2[\text{P}]}$$

where Δ is the observed chemical shift change, Δ_0 is the total change in chemical shift at saturation, and $[\text{L}]$ and $[\text{P}]$ are the peptide and $^4\text{F1}^5\text{F1}$ concentrations, respectively.

Received 20 December 2002; accepted 21 March 2003; doi:10.1038/nature01589.

1. Peacock, S. J., Foster, T. J., Cameron, B. J. & Berendt, A. R. Bacterial fibronectin-binding proteins and endothelial cell surface fibronectin mediate adherence of *Staphylococcus aureus* to resting human endothelial cells. *Microbiology* **145**, 3477–3486 (1999).
2. Ozeri, V., Rosenshine, I., Mosher, D. F., Fässler, R. & Hanski, E. Roles of integrins and fibronectin in the entry of *Streptococcus pyogenes* into cells via protein F1. *Mol. Microbiol.* **30**, 625–637 (1998).
3. Patti, J. M., Allen, B. L., McGavin, M. J. & Höök, M. MSCRAMM-mediated adherence of microorganisms to host tissues. *Annu. Rev. Microbiol.* **48**, 585–617 (1994).
4. House-Pompeo, K., Xu, Y., Joh, D., Speziale, P. & Höök, M. Conformational changes in the fibronectin binding MSCRAMMs are induced by ligand binding. *J. Biol. Chem.* **271**, 1379–1384 (1996).
5. Penkett, C. J. *et al.* Structural and dynamical characterization of a biologically active unfolded fibronectin-binding protein from *Staphylococcus aureus*. *Biochemistry* **37**, 17054–17067 (1998).
6. McGavin, M. J. *et al.* Fibronectin receptors from *Streptococcus dysgalactiae* and *Staphylococcus aureus*—involvement of conserved residues in ligand binding. *J. Biol. Chem.* **268**, 23946–23953 (1993).
7. Joh, D., Speziale, P., Gurusiddappa, S., Manor, J. & Höök, M. Multiple specificities of the staphylococcal and streptococcal fibronectin-binding microbial surface components recognizing adhesive matrix molecules. *Eur. J. Biochem.* **258**, 897–905 (1998).
8. Sinha, B. *et al.* Fibronectin-binding protein acts as *Staphylococcus aureus* invasins via fibronectin bridging to integrin $\alpha_5\beta_1$. *Cell. Microbiol.* **1**, 101–117 (1999).
9. Knodler, L. A., Celli, J. & Finlay, B. B. Pathogenic trickery: Deception of host cell processes. *Nature Rev. Mol. Cell Biol.* **2**, 578–588 (2001).
10. Ing, M. B., Baddour, L. M. & Bayers, S. A. In *The Staphylococci in Human Disease* (eds Crossley, K. B. & Archer, G. L.) 331–354 (Churchill Livingstone, New York, 1997).
11. Greene, C. *et al.* Adhesion properties of mutants of *Staphylococcus aureus* defective in fibronectin-binding proteins and studies on the expression of *fmb* genes. *Mol. Microbiol.* **17**, 1143–1152 (1995).
12. Potts, J. R., Bright, J. R., Bolton, D., Pickford, A. R. & Campbell, I. D. Solution structure of the N-terminal F1 module pair from human fibronectin. *Biochemistry* **38**, 8304–8312 (1999).
13. Jaffe, J., Natanson-Yaron, S., Caparon, M. G. & Hanski, E. Protein F2, a novel fibronectin-binding protein from *Streptococcus pyogenes*, possesses two binding domains. *Mol. Microbiol.* **21**, 373–384 (1996).
14. Penkett, C. J. *et al.* Identification of residues involved in the interaction of *Staphylococcus aureus* fibronectin-binding protein with the $^4\text{F1}^5\text{F1}$ module pair of human fibronectin using heteronuclear NMR spectroscopy. *Biochemistry* **39**, 2887–2893 (2000).
15. Talay, S. R., Valentin-Weigand, P., Jerlstrom, P. G., Timmis, K. N. & Chhatwal, G. S. Fibronectin-binding protein of *Streptococcus pyogenes*—sequence of the binding domain involved in adherence of streptococci to epithelial cells. *Infect. Immun.* **60**, 3837–3844 (1992).
16. Signäs, C. *et al.* Nucleotide sequence of the gene for a fibronectin-binding protein from *Staphylococcus aureus*—use of this peptide sequence in the synthesis of biologically-active peptides. *Proc. Natl. Acad. Sci. USA* **86**, 699–703 (1989).
17. Schwarz-Linek, U. *et al.* Binding of a peptide from a *Streptococcus dysgalactiae* MSCRAMM to the N-terminal F1 module pair of human fibronectin involves both modules. *FEBS Lett.* **497**, 137–140 (2001).
18. Huff, S., Matsuka, Y. V., McGavin, M. J. & Ingham, K. C. Interaction of N-terminal fragments of fibronectin with synthetic and recombinant D motifs from its binding protein on *Staphylococcus aureus* studied using fluorescence anisotropy. *J. Biol. Chem.* **269**, 15563–15570 (1994).
19. Massey, R. C. *et al.* Fibronectin-binding protein A of *Staphylococcus aureus* has multiple, substituting, binding regions that mediate adherence to fibronectin and invasion of endothelial cells. *Cell. Microbiol.* **3**, 839–851 (2001).
20. Talay, S. R. *et al.* Co-operative binding of human fibronectin to SfbI protein triggers streptococcal invasion into respiratory epithelial cells. *Cell. Microbiol.* **2**, 521–535 (2000).
21. Derrick, J. P. & Wigley, D. B. Crystal structure of a streptococcal protein-G domain bound to an Fab fragment. *Nature* **359**, 752–754 (1992).
22. Pickford, A., Smith, S., Staunton, D., Boyd, J. & Campbell, I. The hairpin structure of the $^4\text{F1}^2\text{F2}$ fragment from human fibronectin enhances gelatin binding. *EMBO J.* **20**, 1519–1529 (2001).
23. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**, 289–302 (1999).
24. Brünger, A. T. *X-PLOR (Version 3.1) A System for X-ray Crystallography and NMR* (Yale University, New Haven, Connecticut, 1992).
25. Sass, H. J., Musco, G., Stahl, S. J., Wingfield, P. T. & Grzesiek, S. Solution NMR of proteins within polyacrylamide gels: Diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J. Biomol. NMR* **18**, 303–309 (2000).

26. Ottiger, M., Delaglio, F. & Bax, A. Measurement of J and dipolar couplings from simplified two-dimensional NMR spectra. *J. Magn. Reson.* **131**, 373–378 (1998).
27. Hashimoto, Y. *et al.* The relative orientation of the fibronectin $^6\text{F1}^2\text{F2}$ module pair: A N-15 NMR relaxation study. *J. Biomol. NMR* **17**, 203–214 (2000).
28. Tjandra, N., Garrett, D. S., Gronenborn, A. M., Bax, A. & Clore, G. M. Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. *Nature Struct. Biol.* **4**, 443–449 (1997).
29. Wishart, D. S., Sykes, B. D. & Richards, F. M. The chemical-shift index—a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **31**, 1647–1651 (1992).
30. Talay, S. R., Valentin-Weigand, P., Timmis, K. N. & Chhatwal, G. S. Domain-structure and conserved epitopes of Sfb protein, the fibronectin-binding adhesin of *Streptococcus pyogenes*. *Mol. Microbiol.* **13**, 531–539 (1994).

Acknowledgements We thank R. Aplin for mass spectrometry, M. Pitkeathly for peptide synthesis, and S. Lukowski for the *S. pyogenes* M75 DNA (4673). This research was supported by the Wellcome Trust and the Biotechnology and the Biological Sciences Research Council. J.R.P. acknowledges the British Heart Foundation for financial support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.R.P. (jennifer.potts@bioch.ox.ac.uk). The atomic coordinates have been deposited in the Protein Data Bank with ID code 1O9a.

Nicotinamide and *PNC1* govern lifespan extension by calorie restriction in *Saccharomyces cerevisiae*

Rozalyn M. Anderson*, Kevin J. Bitterman*, Jason G. Wood*, Oliver Medvedik & David A. Sinclair

Department of Pathology, Harvard Medical School, Boston, Massachusetts 02115, USA

* These authors contributed equally to this work

Calorie restriction extends lifespan in a broad range of organisms, from yeasts to mammals. Numerous hypotheses have been proposed to explain this phenomenon, including decreased oxidative damage and altered energy metabolism. In *Saccharomyces cerevisiae*, lifespan extension by calorie restriction requires the NAD^+ -dependent histone deacetylase, Sir2 (ref. 1). We have recently shown that Sir2 and its closest human homologue SIRT1, a p53 deacetylase, are strongly inhibited by the vitamin B_3 precursor nicotinamide². Here we show that increased expression of *PNC1* (pyrazinamidase/nicotinamidase 1), which encodes an enzyme that deaminates nicotinamide, is both necessary and sufficient for lifespan extension by calorie restriction and low-intensity stress. We also identify *PNC1* as a longevity gene that is responsive to all stimuli that extend lifespan. We provide evidence that nicotinamide depletion is sufficient to activate Sir2 and that this is the mechanism by which *PNC1* regulates longevity. We conclude that yeast lifespan extension by calorie restriction is the consequence of an active cellular response to a low-intensity stress and speculate that nicotinamide might regulate critical cellular processes in higher organisms.

Lifespan in the budding yeast *S. cerevisiae* is extended by a variety of stimuli such as heat stress, osmotic stress and the restriction of amino acids or glucose^{1,3–5}. The latter two regimens are considered to be mimics of calorie restriction in higher organisms. In *S. cerevisiae*, replicative age is defined as the number of divisions that a cell undergoes before dying. The yeast *SIR2* gene, which encodes the founding member of a conserved family of NAD^+ -

dependent deacetylases^{6–9}, is required for lifespan extension by glucose restriction¹. Cells with an additional copy of *SIR2* live 30% longer than the wild type, whereas *sir2Δ* strains age prematurely¹⁰ owing to increased recombination at the ribosomal DNA (rDNA) locus^{10,11}. The importance of elucidating the yeast *SIR2* pathway is underscored by increasing evidence that Sir2 proteins in higher organisms promote longevity and cell viability^{12–15}.

Because Sir2 protein levels do not increase in response to calorie restriction¹⁶, lifespan extension must involve an increase in enzymatic activity of Sir2. One hypothesis is that Sir2 is activated by an increased availability of NAD^+ (ref. 1). Nicotinamide, a product of the Sir2 reaction¹⁷, is a strong non-competitive inhibitor of Sir2-like enzymes *in vitro*^{2,17} and can accelerate yeast ageing by inhibiting Sir2 *in vivo*². Thus an alternative explanation is that Sir2 is regulated by changes in nicotinamide levels.

To explore the latter hypothesis, we focused on *PNC1*, a gene whose product converts nicotinamide to nicotinic acid in the NAD^+ salvage pathway (Fig. 1a, b). Most wild-type yeast strains have an average lifespan of 21–23 divisions, with a maximum lifespan of about 40 divisions. A wild-type strain that was calorie restricted (0.5% glucose) or heat stressed (37 °C) exhibited a longer lifespan than an untreated control (2.0% glucose or 30 °C, respectively; Fig. 1c, d). The *sir2Δ* strain had a short lifespan, consistent with previous reports^{1,10}, and neither calorie restriction (0.5% or 0.1% glucose) nor heat stress extended lifespan in this strain (Fig. 1c, d, and data not shown). The *pnc1Δ* strain did not exhibit a lifespan extension under either of these conditions, demonstrating that *PNC1* is necessary for lifespan extension by calorie restriction and low-intensity stress.

Strikingly, under non-stressing conditions (2% glucose, 30 °C), a strain with additional copies of *PNC1* (5×*PNC1*) lived 70% longer than the wild type and some cells lived for more than 70 divisions, which is the longest reported lifespan extension in this organism (Fig. 1e). Neither calorie restriction nor heat stress further increased the lifespan of the 5×*PNC1* strain (not shown). Deletion of *SIR2* in the 5×*PNC1* background shortened lifespan to that of the *sir2Δ* mutant (Fig. 1e). Furthermore, the *pnc1Δ sir2Δ* double mutant had a lifespan similar to that of the *sir2Δ* mutant (Fig. 1e) and its lifespan was unaffected by glucose restriction (not shown). These findings indicate that *PNC1* and *SIR2* function in the same pathway and that *PNC1* increases lifespan through *SIR2*. Together, these results show that *PNC1* is necessary for lifespan extension by both calorie restriction and heat stress, and that additional *PNC1* is sufficient to mimic these stimuli.

Given that additional *PNC1* is sufficient to extend lifespan, we examined whether *PNC1* expression is upregulated in response to stimuli that extend lifespan. We found that Pnc1 levels were greatly induced in a dose-dependent manner by glucose restriction (Fig. 2a) and in cells carrying a *cdc25-10* allele, which mimics calorie restriction¹ (Fig. 2b). *MSN2* and *MSN4*, which encode transcription factors that coordinate the response to carbon source starvation and intense stress, were not required for Pnc1 induction (not shown). This is consistent with the previous observation that these two genes are not required for lifespan extension by glucose restriction¹.

Pnc1 levels were elevated under every other condition known to extend yeast lifespan, including amino acid restriction, salt stress and heat stress (Fig. 2c), in agreement with whole-genome mRNA analyses of stressed yeast cells¹⁸. Pnc1 activity in extracts from treated cells was correlated with Pnc1 concentrations in western blots (Fig. 2d), showing that these cells have increased rates of nicotinamide hydrolysis.

We and others have previously shown that two other enzymes in the NAD^+ salvage pathway, Npt1 (nicotinic acid phosphoribosyltransferase) and Nma2 (nicotinic acid mononucleotide adenylyltransferase), are concentrated in the nucleus^{16,19}. Surprisingly, a fusion protein of Pnc1 with green fluorescent protein (Pnc1-GFP) was not only localized in the nucleus and the cytoplasm but was also

concentrated in three to six discrete cytoplasmic foci per cell (Fig. 3a–d). Calorie-restricted (Fig. 3a) or stressed (Fig. 3b) cells showed a marked increase in the intensity of fluorescence, consistent with the western data. Interestingly, under conditions of amino acid restriction or salt stress, the fluorescence was predominantly localized to the foci (Fig. 3b), suggesting that Pnc1 localization is regulated.

To determine the identity of the foci, we searched for cellular markers that co-localized with Pnc1–GFP and observed significant overlap with a peroxisomally targeted red fluorescent protein (RFP) (Fig. 3c). Pnc1–GFP foci were no longer observed in a peroxisome-deficient *pex6Δ* mutant, confirming that Pnc1–GFP was peroxisomal (Fig. 3d). Because our studies indicated that the localization of Pnc1 to peroxisomes might be regulated, we sought to identify the transporter responsible for its import into this organelle. Although Pex5 imports the vast majority of peroxisomal proteins, the localization of Pnc1 to peroxisomes required the less-used transporter Pex7 (Fig. 3d). The localization of Pnc1 to sites outside

the nucleus implies that this enzyme could regulate proteins other than Sir2 (such as the homologues of Sir2, Hst1–Hst4). The peroxisomal localization of Pnc1 is of particular interest because these organelles are a major source of reactive oxygen species and have been implicated in mammalian ageing²⁰.

Because *PNC1* converts nicotinamide to nicotinic acid as part of the NAD⁺ salvage pathway, it could theoretically activate Sir2 either by increasing the availability of its co-substrate, NAD⁺, or by depleting levels of the inhibitor nicotinamide. Although these mechanisms are not mutually exclusive, and mutations that alter NAD⁺ levels can affect silencing^{1,3,19,21}, current evidence favours the nicotinamide model. We and others have been unable to detect increases in NAD⁺ levels or the NAD⁺/NADH ratio in calorie-restricted cells²² or in genetic mimics of calorie restriction¹⁶, even when unbound NAD⁺ was measured (R.M.A., A. R. Neves, H. Santos and D.A.S., unpublished observations). In addition, we have previously shown that Sir2 is inhibited *in vitro* by physiological concentrations of nicotinamide and that exogenous nicotinamide

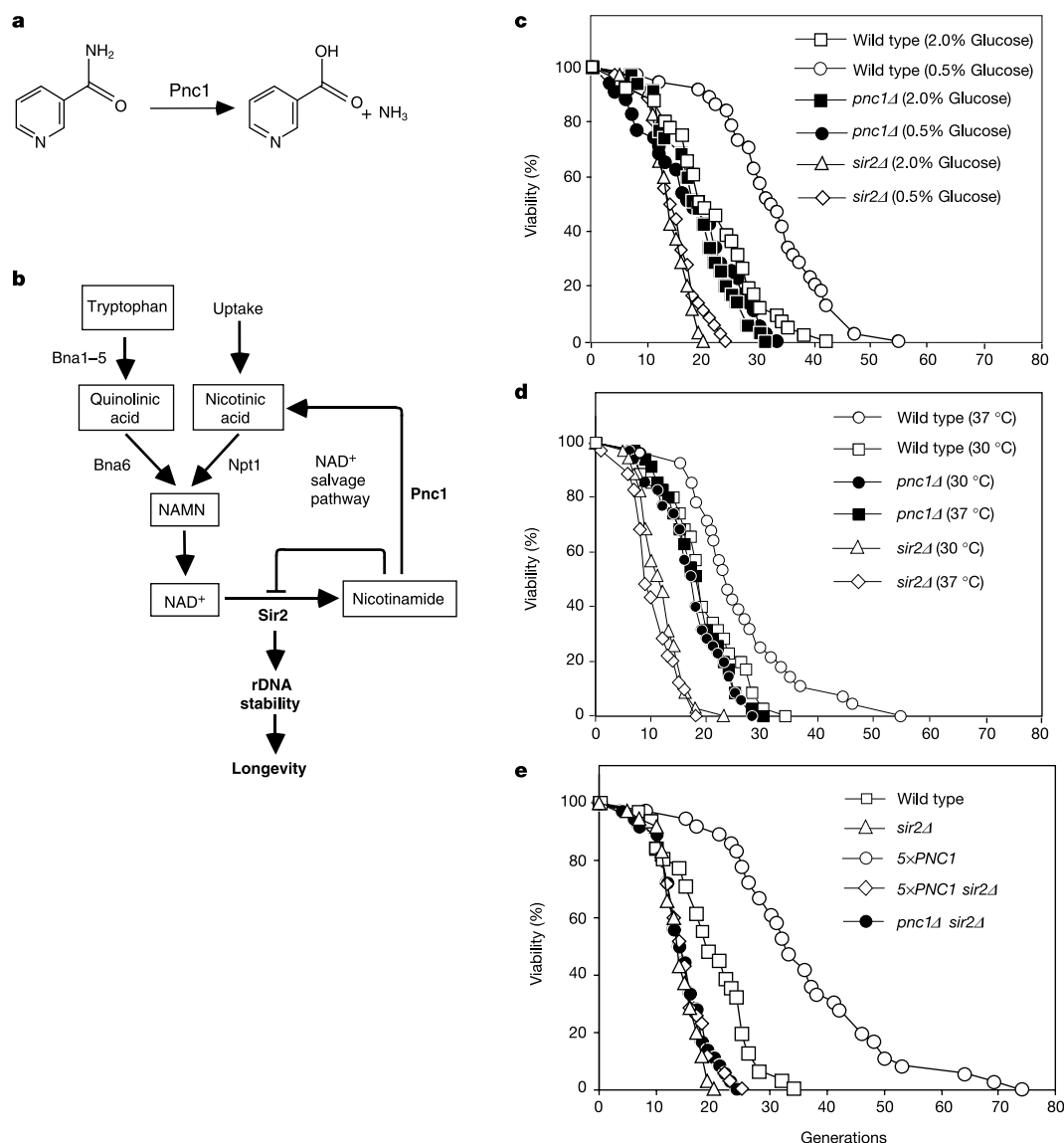


Figure 1 Calorie restriction and heat stress extend lifespan in a *PNC1*-dependent manner. **a**, Pnc1 converts nicotinamide to nicotinic acid. **b**, In *S. cerevisiae*, NAD⁺ is synthesized *de novo* from tryptophan via Bna1–6 or recycled from nicotinamide. **c**, Average lifespan on 2.0% (w/v) glucose: wild type, 21.6 generations; *pnc1Δ*, 19.1; *sir2Δ*, 14.2. Average lifespan on 0.5% glucose: wild type, 32.7 generations; *pnc1Δ*, 18.1; *sir2Δ*, 14.7. **d**, At

30 °C, average lifespans: wild type, 19.4 generations; *pnc1Δ*, 18.5; *sir2Δ*, 12.0. At 37 °C, average lifespans: wild type, 23.4 generations; *pnc1Δ*, 17.5; *sir2Δ*, 10.6.

e, Average lifespans on 2.0% glucose at 30 °C: wild type, 19.7 generations; 5xPNC1, 36.1; *sir2Δ*, 14.2; 5xPNC1 *sir2Δ*, 15.1; *pnc1Δ* *sir2Δ*, 14.4.

can abolish silencing *in vivo*². Perhaps most persuasive is the observation that cells lacking *PNC1* have a silencing defect, yet they show no change in NAD⁺ levels¹⁹. Although these observations are supportive of the nicotinamide model, we sought more conclusive evidence.

First, we reasoned that if Pnc1 activates Sir2 by stimulating the NAD⁺ salvage pathway (by converting nicotinamide to nicotinic acid), then an increase in the intracellular nicotinic acid pool should have the same effect as increasing Pnc1 levels (see Fig. 1b). Exogenous nicotinic acid is readily taken up by yeast cells and can significantly increase the intracellular pool (R.M.A., A. R. Neves, H. Santos and D.A.S., unpublished observations)²³. A common indicator of Sir2 activity is the extent to which a reporter gene inserted at the rDNA locus (*RDN1*) is silenced. As shown in Fig. 4a, exogenous nicotinic acid did not increase rDNA silencing, indicating that nicotinic acid is not limiting for the salvage pathway. Furthermore, genetic analysis demonstrates that the contribution of *PNC1* to NAD⁺ synthesis is minimal, even in the absence of NAD⁺ synthesis *de novo* (Fig. 4b). Taken together, these data argue against a model in which Pnc1 stimulates Sir2 by providing additional nicotinic acid for NAD⁺ synthesis.

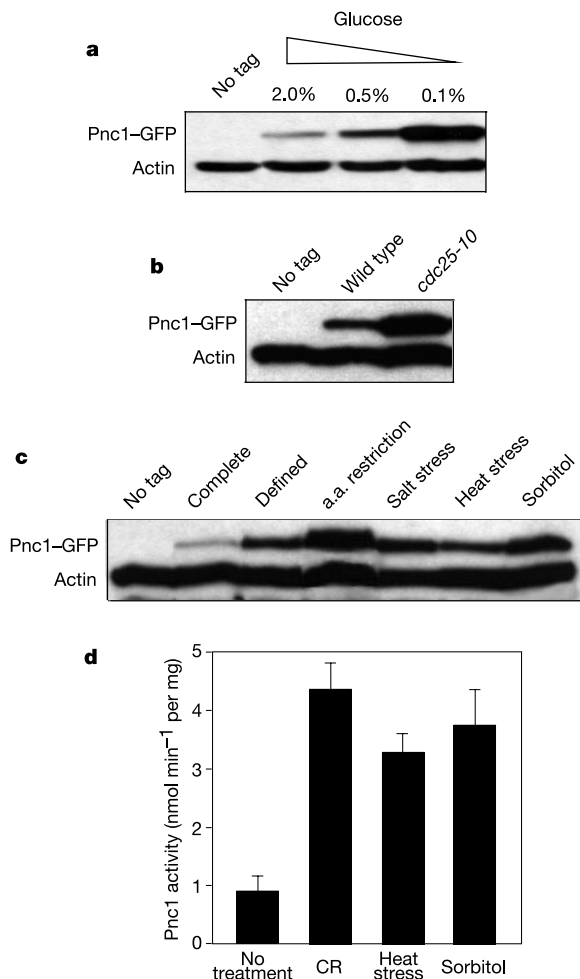


Figure 2 Pnc1 levels and activity are elevated in response to calorie restriction and low-intensity stress. **a**, Western analysis of Pnc1-GFP under conditions of glucose restriction. **b**, Pnc1-GFP in wild-type or *cdc25-10* cells. **c**, Detection of Pnc1-GFP in cells subjected to mild stress as indicated (a.a., amino acid). **d**, Measurement of nicotinamide deamination by Pnc1. Activity (nmol ammonia min⁻¹ per mg protein) from three experiments (means $\pm$ s.d.): no treatment (2% glucose), 0.9 ± 0.26 ; calorie restriction (CR; 0.1% glucose), 4.38 ± 0.43 ; heat stress (37 °C), 3.28 ± 0.32 ; sorbitol (1 M), 3.75 ± 0.65 .

Second, we tested whether the manipulation of *PNC1* could increase silencing even when its contribution to NAD⁺ synthesis was blocked. In *S. cerevisiae*, the only other NAD⁺ salvage pathway gene that can be deleted without a loss of viability is *NPT1* (see Fig. 1b). We have previously shown that additional copies of *PNC1* increase rDNA silencing in wild-type cells¹⁶. Additional copies of *PNC1* led to a partial rescue of the silencing defect in the *npt1Δ* strain (Fig. 4c). Because cells lacking *NPT1* have NAD⁺ levels one-half of those in wild-type strains⁶, we included an NAD⁺ precursor, quinolinic acid, in the medium, which in mammals has been shown to compensate for a low NAD⁺ concentration²⁴. In the presence of this compound, additional *PNC1* restored rDNA silencing in the *npt1Δ* strain to near wild-type levels (Fig. 4c), showing that Pnc1 can increase Sir2 activity even in the absence of the NAD⁺ salvage pathway.

Last, if *PNC1* regulates Sir2 activity by modulating nicotinamide levels, we reasoned that manipulation of nicotinamide using a gene outside the NAD⁺ salvage pathway should have the same effect. In humans, nicotinamide is converted to *N*-methylnicotinamide by nicotinamide *N*-methyltransferase (NNMT)²⁵ and then excreted. As predicted by the nicotinamide model, overexpression of human NNMT in yeast increased rDNA silencing (Fig. 4d). By homology we also identified a putative *S. cerevisiae* NNMT gene, YLR285W. The predicted protein contains the four signature motifs of *S*-adenosylmethionine-dependent methyltransferases²⁶ and its core domain is 23% identical to that of human NNMT²⁵. Additional copies of YLR285W increased silencing, whereas deletion of this gene led to a loss of silencing, similar to the effect of manipulating *PNC1* (ref. 16) (Fig. 4d). Additional copies of YLR285W increased

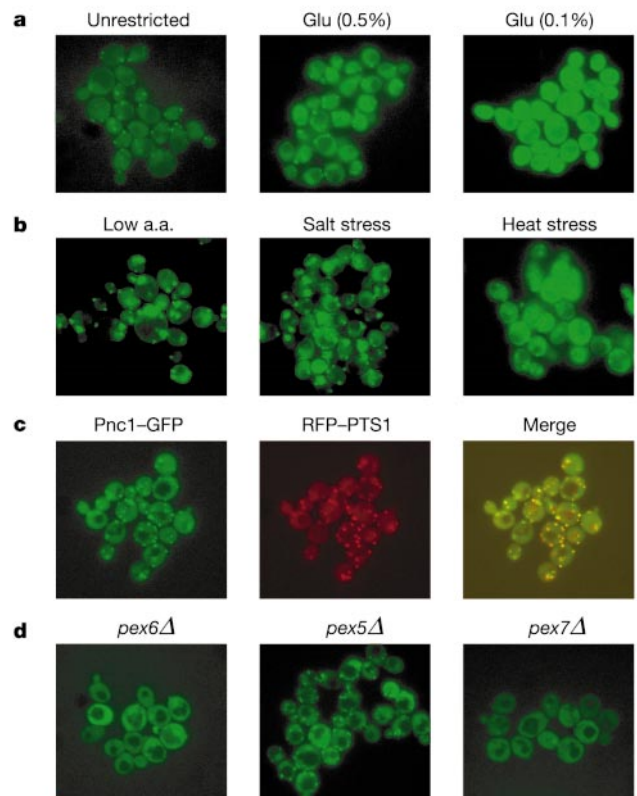


Figure 3 Pnc1-GFP is localized in the nucleus and cytoplasm, and concentrated in peroxisomes. **a**, Pnc1-GFP fluorescence in glucose-restricted cells (Glu 0.5% and 0.1%). **b**, Pnc1-GFP fluorescence under conditions of mild stress (a.a., amino acid). **c**, Co-localization of Pnc1-GFP (green) and RFP-PTS1 (red). Yellow indicates overlap. **d**, Localization of Pnc1-GFP in cells from peroxisomal mutant strains, *pex6Δ*, *pex5Δ* and *pex7Δ*.

yeast lifespan and this effect was not enhanced by glucose restriction (Fig. 4e). Unlike *PNC1*, YLR285W is not a true longevity regulator because its expression is not apparently modulated by stimuli that extend lifespan¹⁸, and its deletion does not abolish lifespan extension by glucose restriction (Fig. 4e).

Our results show that lifespan extension by either calorie restriction or mild stress is the result of an active cellular response that requires the upregulation of a specific longevity gene, *PNC1* (Fig. 4f). This system of longevity regulation explains how multiple, disparate stimuli can lead to the same longevity response and how species can rapidly evolve strategies to suit a changing environment. We also provide multiple lines of evidence that *PNC1* regulates Sir2 by modulating intracellular nicotinamide. It has been proposed that Sir2 is regulated by passive means, by changes in either NAD⁺ availability^{1,3,6,21} or the NAD⁺/NADH ratio^{1,3,21}. We do not exclude

the possibility that these mechanisms can function in tandem with nicotinamide depletion. However, an attractive feature of nicotinamide-based regulation is that it does not require the modulation of NAD⁺, an essential cofactor involved in cellular homeostasis.

Nicotinamide has been shown to promote apoptosis in mammalian cells by inhibiting the Sir2 homologue SIRT1 (refs 2, 15), a regulator of p53 (refs 14, 15). Moreover, the poly(ADP-ribose) polymerase family of proteins, which are involved in many processes including DNA repair, telomere-length regulation and the opening of chromatin associated with stress-activated genes, are also inhibited by nicotinamide²⁷. Interestingly, an increased expression of *NNMT* is correlated with tumorigenesis²⁸ and a decreased expression is correlated with radiosensitivity²⁹. These findings raise the possibility that nicotinamide regulates critical cellular processes in higher organisms. □

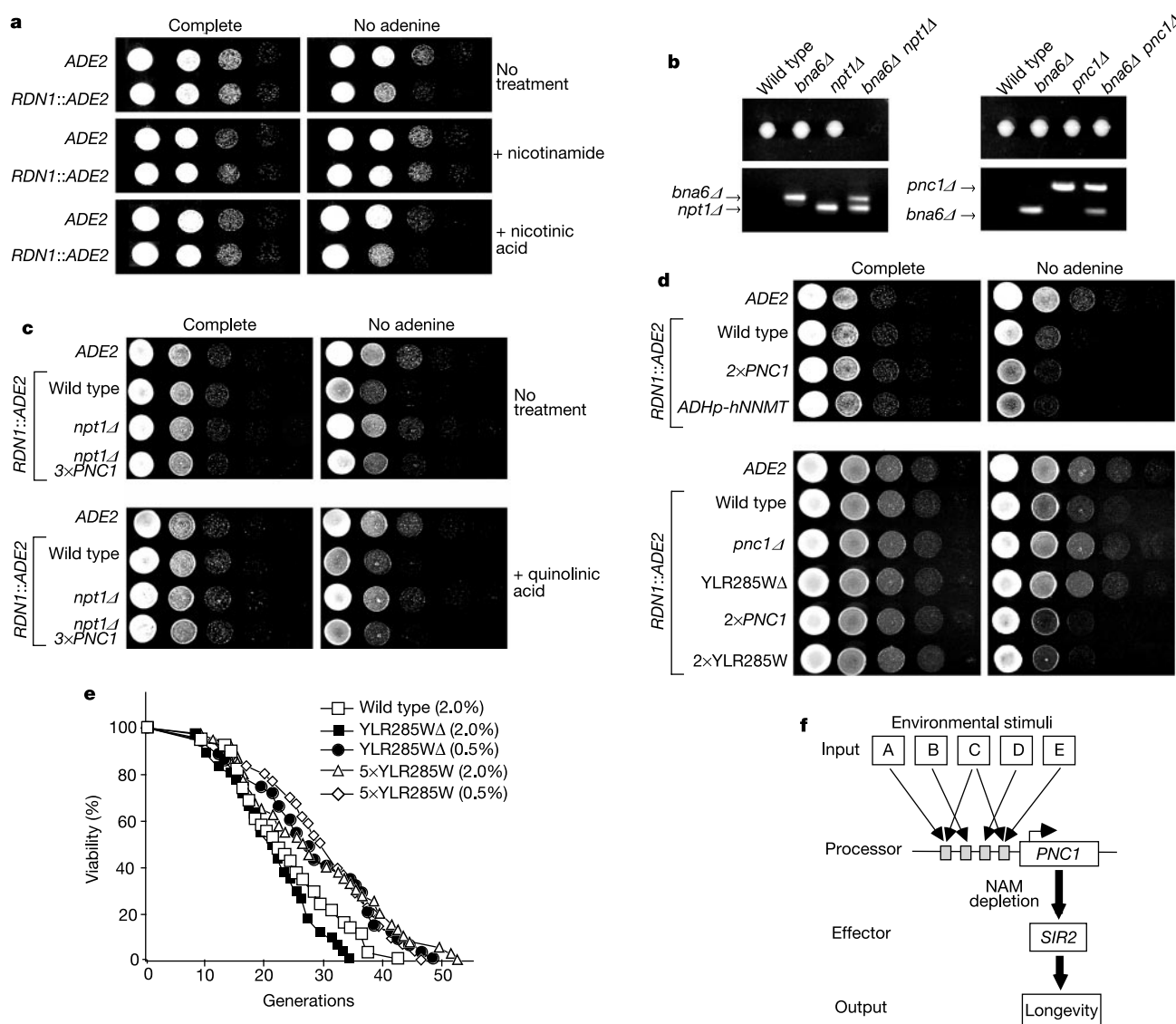


Figure 4 Manipulation of nicotinamide metabolism alters silencing and lifespan.

a, To monitor silencing, *ADE2* was integrated at the rDNA locus. Increased growth on medium lacking adenine indicates decreased silencing. Serial dilutions of cells spotted on plates containing nicotinic acid or nicotinamide (5 mM). **b**, *PNC1* does not have a critical role in NAD⁺ biosynthesis, even in the absence of the *de novo* NAD⁺ synthesis pathway. *BNAF6* encodes an enzyme in the NAD⁺ *de novo* synthesis pathway (see Fig. 1b). *NPT1* encodes a phosphoribosyltransferase that converts nicotinic acid to nicotinamide

mononucleotide in the NAD⁺ salvage pathway. Spore colonies from heterozygous *bnaf6Δ npt1Δ* or *bnaf6Δ pnc1Δ* diploids. Genotypes determined by PCR with a colony/microcolony genomic template. **c**, Partial rescue of silencing by additional *PNC1* in the absence of the NAD⁺ salvage pathway and complete rescue in the presence of quinolinic acid (5 mM). **d**, Manipulation of genes involved in nicotinamide metabolism alters rDNA silencing. **e**, Manipulation of YLR285W affects lifespan. **f**, Model for the regulation of Sir2 activity and lifespan by *PNC1* and nicotinamide (NAM).

Methods

Media and strains

All strains were grown at 30 °C in complete 2.0% (w/v) glucose (YPD) medium except where stated otherwise. Glucose restriction medium contained 0.5% or 0.1% glucose. Mild stress conditions were one of the following: defined medium (SD); amino acid restriction (SD lacking non-essential amino acids); salt stress (NaCl, 300 mM); heat stress (37 °C); sorbitol (1 M). In all experiments, auxotrophic markers were matched between strains by integrating empty vector. The copy number of integrated genes was determined by Southern blotting. Deletions were generated with a kan-MX6 PCR-based technique¹⁶ and confirmed by PCR. Additional copies of *PNC1* were integrated as described previously¹⁶. The entire open reading frame and 700 bases of the upstream sequence of YLR285W were amplified from genomic DNA and cloned into pSP400, then sequenced and integrated as described previously¹⁶. A GFP cassette was integrated in frame at the 3' end of the native *PNC1* gene as described previously¹⁶. The RFP-PTS1 (for peroxisomal targeting signal 1) plasmid (pSG421) was a gift from S. J. Gould (Johns Hopkins University). The coding region of human *NNMT* was subcloned from p91023(B), a gift from R. Weinshilboum (Mayo Clinic), into pSP400 downstream of the *ADH1* promoter. Strains derived from PSY316AT¹⁶ were used for lifespan analysis. Strains derived from W303AR¹¹ were used for western blots, microscopy and silencing assays. W303AR *cdc25-10* was a gift from L. Guarente (MIT).

Yeast assays

Lifespan measurements were performed as described previously¹⁶ except for the heat-stress experiments, in which strains were incubated after each dissection at 37 °C. Silencing was assayed as described previously¹⁶.

Protein expression analysis

Strains were pretreated under the indicated conditions and grown to mid-exponential phase. Western blots were performed as described¹⁶ with whole-cell extracts (75 µg). Proteins were detected with anti-GFP antibodies (Santa Cruz) or anti-actin antibodies (Chemicon). Fluorescent microscopy images were captured at the same exposure (1 s) at ×100 magnification with a Hamamatsu Orca100 CCD camera and processed with Openlab software. Cultures were grown in complete medium containing 0.5% glucose to enhance the detection of fluorescence.

Nicotinamidase activity assay

The activity of Pnc1 in extracts obtained from pretreated mid-exponential-phase cultures was determined as described previously³⁰. In brief, 0.16 mg of protein was incubated with either 0 or 8 mM nicotinamide for 45 min at 30 °C in a final volume of 400 µl of 10 mM Tris-HCl pH 7.5, 150 mM NaCl and 1 mM MgCl₂. Pnc1 activity was determined by measuring the final concentration of the reaction product, ammonia, with an ammonia diagnostic kit (Sigma). Baseline ammonia was accounted for by subtracting a no-nicotinamide control. Nicotinamidase activity was expressed as nmol ammonia min⁻¹ per mg total protein. Pnc1 activity was obtained by subtracting the background value for the *pnc1Δ* strain (0.06 ± 0.004 nmol min⁻¹ per mg).

Received 11 February; accepted 20 March 2003; doi:10.1038/nature01578.

- Lin, S. J., Defossez, P. A. & Guarente, L. Requirement of NAD and *SIR2* for life-span extension by calorie restriction in *Saccharomyces cerevisiae*. *Science* **289**, 2126–2128 (2000).
- Bitterman, K. J., Anderson, R. M., Cohen, H. Y., Latorre-Esteves, M. & Sinclair, D. A. Inhibition of silencing and accelerated aging by nicotinamide, a putative negative regulator of yeast Sir2 and human SIRT1. *J. Biol. Chem.* **277**, 45099–45107 (2002).
- Kaeberlein, M., Andalis, A. A., Fink, G. R. & Guarente, L. High osmolarity extends life span in *Saccharomyces cerevisiae* by a mechanism related to calorie restriction. *Mol. Cell. Biol.* **22**, 8056–8066 (2002).
- Jiang, J. C., Jaruga, E., Repnevskaya, M. V. & Jazwinski, S. M. An intervention resembling caloric restriction prolongs life span and retards aging in yeast. *FASEB J.* **14**, 2135–2137 (2000).
- Swiciclo, A., Krawiec, Z., Wawryn, J., Bartosz, G. & Bilinski, T. Effect of stress on the life span of the yeast *Saccharomyces cerevisiae*. *Acta Biochim. Pol.* **47**, 355–364 (2000).
- Smith, J. S. *et al.* A phylogenetically conserved NAD⁺-dependent protein deacetylase activity in the Sir2 protein family. *Proc. Natl Acad. Sci. USA* **97**, 6658–6663 (2000).
- Imai, S., Armstrong, C. M., Kaeberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
- Tanny, J. C. & Moazed, D. Coupling of histone deacetylation to NAD breakdown by the yeast silencing protein Sir2: Evidence for acetyl transfer from substrate to an NAD breakdown product. *Proc. Natl Acad. Sci. USA* **98**, 415–420 (2001).
- Landry, J. *et al.* The silencing protein Sir2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl Acad. Sci. USA* **97**, 5807–5811 (2000).
- Kaeberlein, M., McVey, M. & Guarente, L. The Sir2/3/4 complex and Sir2 alone promote longevity in *Saccharomyces cerevisiae* by two different mechanisms. *Genes Dev.* **13**, 2570–2580 (1999).
- Sinclair, D. A. & Guarente, L. Extrachromosomal rDNA circles—a cause of aging in yeast. *Cell* **91**, 1033–1042 (1997).
- Tissenbaum, H. A. & Guarente, L. Increased dosage of a *sir-2* gene extends lifespan in *Caenorhabditis elegans*. *Nature* **410**, 227–230 (2001).
- Rogina, B., Helfand, S. L. & Frankel, S. Longevity regulation by *Drosophila* Rpd3 deacetylase and caloric restriction. *Science* **298**, 1745 (2002).
- Vaziri, H. *et al.* hSIR2(SIRT1) Functions as an NAD-dependent p53 deacetylase. *Cell* **107**, 149–159 (2001).
- Luo, J. *et al.* Negative control of p53 by Sir2α promotes cell survival under stress. *Cell* **107**, 137–148 (2001).
- Anderson, R. M. *et al.* Manipulation of a nuclear NAD⁺ salvage pathway delays aging without altering steady-state NAD⁺ levels. *J. Biol. Chem.* **277**, 18881–18890 (2002).
- Landry, J., Slama, J. T. & Sternglanz, R. Role of NAD⁺ in the deacetylase activity of the SIR2-like

proteins. *Biochem. Biophys. Res. Commun.* **278**, 685–690 (2000).

- Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell* **11**, 4241–4257 (2000).
- Sandmeier, J. J., Celic, I., Boeke, J. D. & Smith, J. S. Telomeric and rDNA silencing in *Saccharomyces cerevisiae* are dependent on a nuclear NAD⁺ salvage pathway. *Genetics* **160**, 877–889 (2002).
- Perichon, R., Bourre, J. M., Kelly, J. F. & Roth, G. S. The role of peroxisomes in aging. *Cell Mol. Life Sci.* **54**, 641–652 (1998).
- Lin, S. J. *et al.* Calorie restriction extends *Saccharomyces cerevisiae* lifespan by increasing respiration. *Nature* **418**, 344–348 (2002).
- Lin, S. S., Manchester, J. K. & Gordon, J. I. Enhanced gluconeogenesis and increased energy storage as hallmarks of aging in *Saccharomyces cerevisiae*. *J. Biol. Chem.* **276**, 36000–36007 (2001).
- Unkefer, C. J. & London, R. E. *In vivo* studies of pyridine nucleotide metabolism in *Escherichia coli* and *Saccharomyces cerevisiae* by carbon-13 NMR spectroscopy. *J. Biol. Chem.* **259**, 2311–2320 (1984).
- Grant, R. S. & Kapoor, V. Murine glial cells regenerate NAD, after peroxide-induced depletion, using either nicotinic acid, nicotinamide, or quinolinic acid as substrates. *J. Neurochem.* **70**, 1759–1763 (1998).
- Aksoy, S., Szumlanski, C. L. & Weinshilboum, R. M. Human liver nicotinamide N-methyltransferase. cDNA cloning, expression, and biochemical characterization. *J. Biol. Chem.* **269**, 14835–14840 (1994).
- Niewmierzyczna, A. & Clarke, S. S-Adenosylmethionine-dependent methylation in *Saccharomyces cerevisiae*. Identification of a novel protein arginine methyltransferase. *J. Biol. Chem.* **274**, 814–824 (1999).
- Virag, L. & Szabo, C. The therapeutic potential of poly(ADP-ribose) polymerase inhibitors. *Pharmacol. Rev.* **54**, 375–429 (2002).
- Lal, A. *et al.* A public database for gene expression in human cancers. *Cancer Res.* **59**, 5403–5407 (1999).
- Kassem, H., Sangar, V., Cowan, R., Clarke, N. & Margison, G. P. A potential role of heat shock proteins and nicotinamide N-methyl transferase in predicting response to radiation in bladder cancer. *Int. J. Cancer* **101**, 454–460 (2002).
- Ghislain, M., Talla, E. & Francois, J. M. Identification and functional analysis of the *Saccharomyces cerevisiae* nicotinamidase gene, *PNC1*. *Yeast* **19**, 215–224 (2002).

Acknowledgements We thank members of the Sinclair laboratory, R. Veech, C. Wolberger, W. Forrester, S. Luikenhuis and D. Finkelstein, for reagents and discussions. This work was supported by the NIA and the Harvard–Armenise Foundation. D.S. is an Ellison Medical Research Foundation Special Fellow. R.A. is supported by a John Taplan Postdoctoral Fellowship, J.W. by a National Science Foundation Scholarship, and K.B. and O.M. by the American Federation of Aging Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.A.S. (david_sinclair@hms.harvard.edu).

Computational design of receptor and sensor proteins with novel functions

Loren L. Looger, Mary A. Dwyer, James J. Smith & Homme W. Hellinga

Department of Biochemistry, Duke University Medical Center, Durham, North Carolina 27710, USA

The formation of complexes between proteins and ligands is fundamental to biological processes at the molecular level. Manipulation of molecular recognition between ligands and proteins is therefore important for basic biological studies¹ and has many biotechnological applications, including the construction of enzymes^{2–4}, biosensors^{5,6}, genetic circuits⁷, signal transduction pathways⁸ and chiral separations⁹. The systematic manipulation of binding sites remains a major challenge. Computational design offers enormous generality for engineering protein structure and function¹⁰. Here we present a structure-based computational method that can drastically redesign protein ligand-binding specificities. This method was used to construct soluble receptors that bind trinitrotoluene, L-lactate or serotonin with high selectivity and affinity. These engineered receptors can function as biosensors for their new ligands; we also incorporated them into synthetic bacterial signal transduction pathways, regulating gene expression in response to extracellular trinitrotoluene or L-lactate. The use of various ligands

and proteins shows that a high degree of control over biomolecular recognition has been established computationally. The biological and biosensing activities of the designed receptors illustrate potential applications of computational design.

The general principles for the formation of specific complexes are well understood¹¹, and involve a lock-and-key fit between ligand and receptor, the structure of which is determined primarily by short-range interactions (steric contacts and hydrogen bonds). Complex formation is thermodynamically driven primarily by the hydrophobic effect and by long-range electrostatics. Difficulties in structure-based computational design arise from limitations in the description of the molecular interactions and the combinatorial complexity of the problem^{12,13}. Despite notable advances in the rational manipulation of protein sequence and stability using automated computational design tools¹², the computational design of ligand-binding properties has been limited to metal centres⁴, changes in specificity in which much of the chemical character of the wild-type ligand is retained^{2,8,14} or has resulted in relatively weak binding³.

Here we present a computational method for the redesign of

ligand-binding-site specificity in proteins (see Supplementary Information for details). By using high-resolution three-dimensional structures, the algorithm identifies amino-acid sequences that are predicted to form a complementary surface between the protein and a target ligand replacing the wild-type ligand. The procedure combines target-ligand docking ($\sim 10^8$ translations and rotations) with mutations of amino-acid residues in direct contact with the wild-type ligand (typically 12–18 residues, corresponding to 10^{45} to 10^{68} mutant structures representing 10^{15} to 10^{23} sequences). The resulting combinatorial problem (10^{53} to 10^{76} choices) is solved with an algorithm based on the dead-end elimination (DEE) theorems¹³. This procedure deterministically identifies the global minimum of a semi-empirical potential function describing the molecular interactions in the system¹³, including a modified Lennard-Jones potential, an explicit, geometry-dependent hydrogen-bonding term and a continuum solvation term to represent the hydrophobic effect. Additionally, a new term demanding that potential hydrogen-bond donors and acceptors in the ligand must be satisfied was found to be critical; it captures the necessity of balancing the hydrogen bond inventory, which is a dominant effect in molecular recognition¹¹. Designs are selected for experimentation from a rank-ordered set of possibilities. The design process is relatively rapid, requiring about 3 days of computation to generate a set of designs in a particular protein for a given ligand on a 20-processor computer cluster.

This procedure was used to engineer binding sites for trinitrotoluene (TNT), L-lactate or serotonin in place of the wild-type sugar or amino-acid ligands of five members of the *Escherichia coli* periplasmic binding protein (PBP) superfamily¹⁵, using the high-resolution three-dimensional structures of the closed conformation of these proteins complexed with their wild-type ligand as starting points for the calculation (Fig. 1). The five proteins were glucose-binding protein (GBP)¹⁶, ribose-binding protein (RBP)¹⁷, arabinose-binding protein (ABP)¹⁸, glutamine-binding protein (QBP)¹⁹ and histidine-binding protein (HBP)²⁰. The variation in structure and sequence¹⁵ of these proteins presents distinct starting points for the design calculations. The three target ligands selected for this study bear little resemblance to the wild-type cognate ligands of the chosen PBPs; they are chemically distinct from each other and one of them (TNT) is a non-natural molecule. The designs therefore explore critical parameters of molecular recognition, including molecular shape, chirality, functional groups (hydrogen bonding, in terms of nitro (acceptor), hydroxyl (donor and acceptor) and carboxylate (acceptor); and molecular surface, in terms of polar, aliphatic and aromatic), internal flexibility (TNT < L-lactate < serotonin), charge (neutral (TNT), anionic (L-lactate) and cationic (serotonin)) and water solubility (TNT < serotonin < L-lactate).

Complementary surfaces were designed for TNT in RBP, ABP and HBP, for L-lactate in ABP, GBP, RBP, HBP and QBP, and for serotonin in ABP (Fig. 2, Table 1). The designed surfaces are electrically neutral for TNT, positively charged for lactate, and negatively charged for serotonin. Hydrophobic groups of all three target ligands interact primarily with aliphatic side chains, although several examples of aromatic interactions are seen (TNT.A1, TNT.H1, Lac.G1). In one instance, an example of dual aromatic stacking was obtained (TNT.R3). In all cases, the hydrogen-bonding potential (donor, acceptor) of the functional groups on the ligand is largely satisfied.

Seventeen designs predicted by the automated design procedure were selected for experimental characterization (Table 1). The predicted mutations (ranging from 5 to 17 amino-acid changes) were constructed by polymerase chain reaction mutagenesis of the wild-type receptor scaffold genes⁶. Proteins were overexpressed, purified and modified with thiol-reactive styryl dyes conjugated to cysteine residues introduced by mutation at locations where the fluorescence emission intensity of the dye responds to a ligand-mediated hinge-bending motion of the receptor⁶. Ligand-binding

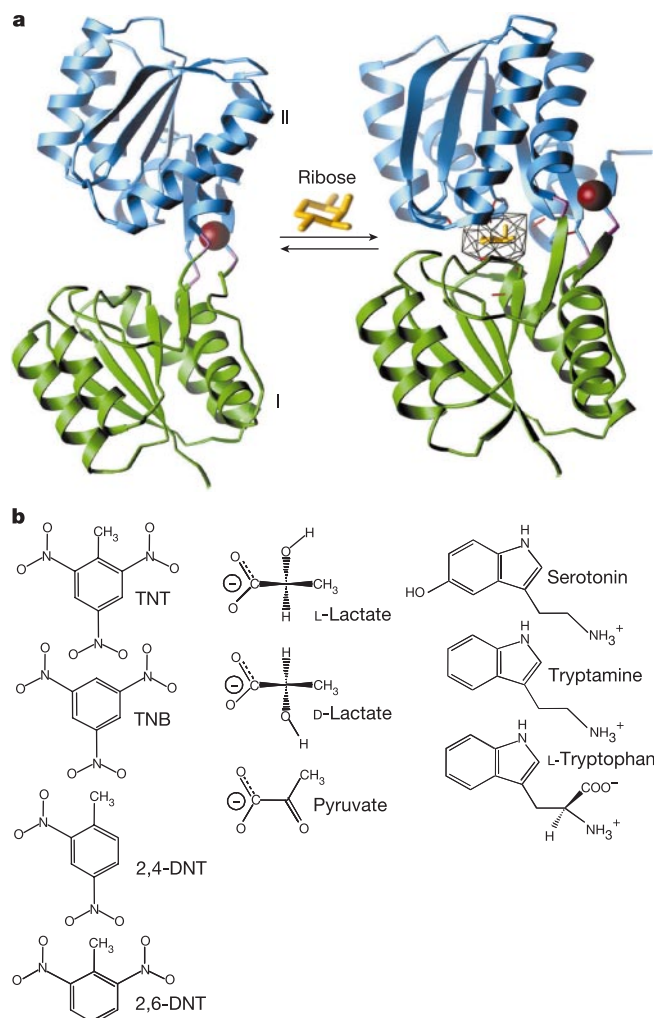


Figure 1 Structures of a representative receptor and all target ligands. **a**, Ribose binding protein mediates a transition from an open (left) to a closed (right) conformation¹⁷. The protein has two domains (I, amino terminal, green; II, carboxy terminal, blue), linked by a hinge region (H, purple). Fluorescence intensity changes of an environmentally sensitive, thiol-reactive fluorescent dye coupled to a mutant cysteine residue at position 265 (burgundy) monitor ligand binding⁶. Calculations use the closed structure, mutating the complementary surface residues (red) and docking the target ligands into the convex hull shown (black). **b**, Structures of target ligands and structurally related decoys used to probe the specificity of the designed receptors.

affinities (Table 2) were determined by titration, monitoring ligand-dependent changes in fluorescence emission intensity that were fitted to single-site binding isotherms⁶ (Supplementary Information). In all cases, wild-type receptors show no change in fluorescence intensity upon addition of target ligands. Conversely, the mutant receptors respond only to target and not wild-type ligands. A wide range of affinities is observed, down to the nanomolar level (TNT.R3). To probe the specificity of interaction, the affinities of several closely related ligands (Fig. 1b) were also determined (Table 2). The thermostabilities of a representative subset of apo-receptors (Supplementary Information) showed that cooperative folding transitions are retained with a slight loss of stability relative to the wild-type proteins.

Every designed receptor exhibits a detectable affinity for its target ligand. In the TNT designs, all six receptors can distinguish the absence of a single nitro group (2,4- and 2,6-dinitrotoluene) and, with the exception of the ABP design, the absence of a single methyl group (trinitrobenzene). The introduction of an additional point mutation suggested by visual inspection of the model to improve packing is sufficient to achieve the desired selectivity in this ABP design. All 10 lactate designs exhibit the desired chiral stereospecificity, selecting L-lactate over both the D-lactate enantiomer and pyruvate, the prochiral oxidized form of lactate. The single serotonin design has a significantly lower affinity for tryptamine (absence of a hydroxyl group) and tryptophan (absence of a hydroxyl group and presence of a carboxylate group). The relative free energy

corresponding to the loss of a hydrogen bond in a decoy ligand ($1\text{--}5\text{ kcal mol}^{-1}$; Supplementary Information) is consistent with the observed range of weak and strong hydrogen bonds¹¹. The automated computational design procedure therefore reliably predicts mutant receptors that attain ligand binding with the desired, drastically altered specificity, consistent with correct modelling of critical elements of molecular recognition: shape, functional groups and chirality.

The affinities of the wild-type receptors for their cognate ligands fall in the range $0.1\text{--}1.5\text{ }\mu\text{M}$ (ref. 6). Two of the three TNT designs in RBP also fall into this range (Table 2); the binding behaviour of these computationally designed receptors is therefore indistinguishable from naturally evolved PBPs. It has been observed that the maximal binding affinity for many ligands is correlated with the number of non-hydrogen atoms²¹. The affinity of one TNT design, TNT.R3, is 2 nM , corresponding to its empirically expected value. The single serotonin design does not attain the expected nanomolar affinity. The affinity of the fully automated design (Stn.A1) is $50\text{ }\mu\text{M}$, and is improved to $4.7\text{ }\mu\text{M}$ by the introduction of a single point mutation predicted to improve packing interactions between receptor and ligand. Several of the lactate designs have micromolar affinities, approaching the expected maximum for a six-atom ligand ($0.3\text{ }\mu\text{M}$).

High-affinity receptors are successfully identified within the top 10 ranked designs for each ligand, corresponding to a tiny fraction of the available search space. Nevertheless, the designs exhibit a significant spread in ligand-binding affinities, not only for a given

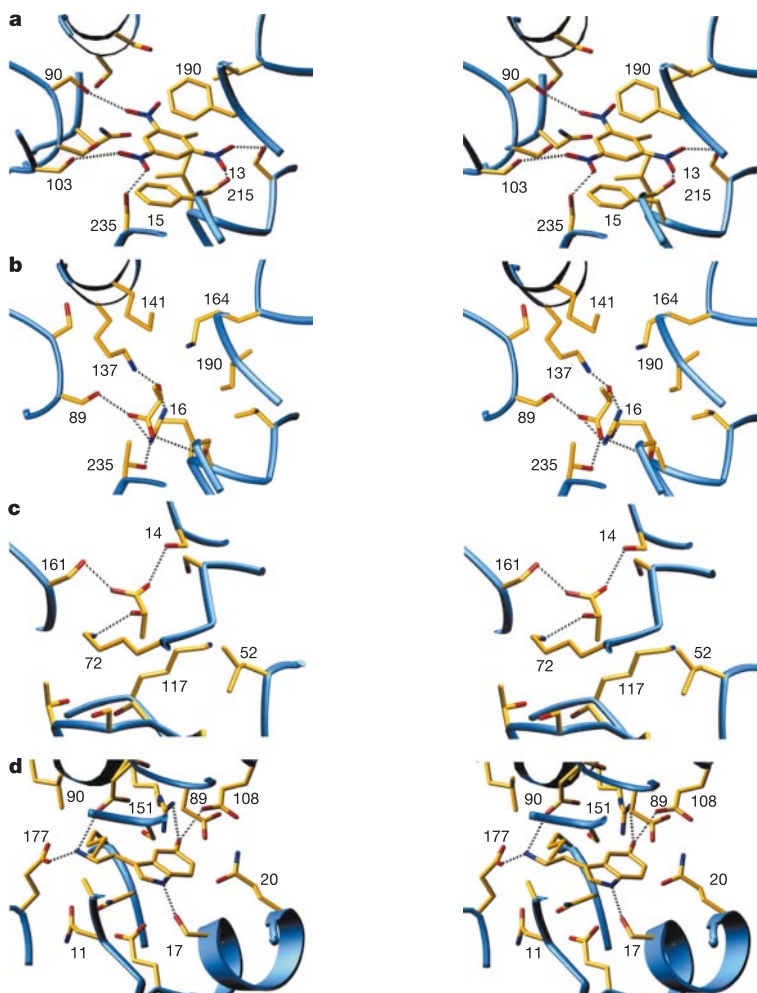


Figure 2 Stereo views of representative designed ligand-binding sites: **a**, TNT.R3; **b**, Lac.R1; **c**, Lac.H1; **d**, Stn.A1 (dashed lines indicate hydrogen bonds between protein and ligand; numbers identify side chains close to the ligand). TNT.R3 and Lac.R1 are

presented in the same orientation, illustrating the adaptability of the RBP scaffold to bind different ligands. The Lac.R1 and Lac.H1 structures illustrate that the same ligand can be bound by sites designed in different scaffolds.

Table 1 Complementary surface sequences of the designed receptors

Scaffold	Target	Design	Complementary surface sequence†																	
RBP			9 _I	13 _I	15 _I	16 _I	64 _I	89 _I	90 _I	103 _H	132 _H	137 _H	141 _{II}	164 _{II}	190 _{II}	214 _{II}	215 _{II}	235 _{II}		
	TNT	Wild-type R1 (I) R2 (I) R3 (A)	S S S	N N I	F A F	F N L	N S S	D S A S S	R R D S S	S	I K	A	R R A S M	F S N I K	N N N F I	F K	D A A S S	Q I S I S I I		
	Lac	R1 (A)	S V	S	A	R	S	S	S	S	S	S K	S	M	K	F	S	I		
GBP			10 _I	14 _I	16 _I	91 _I	92 _I	152 _{II}	154 _{II}	158 _{II}	183 _{II}	211 _{II}	236 _{II}	256 _{II}						
	Lac	Wild-type G1(A) G2(A)	Y K K	D K M	F F K	N K K	K L L	H M K	D H K	R K M	W K K	N N N	D A A	N D S						
ABP			10 _I	11 _I	14 _I	16 _I	17 _I	20 _I	64 _I	89 _I	90 _I	108 _{II}	145 _{II}	147 _{II}	151 _{II}	177 _{II}	204 _{II}	205 _{II}	232 _{II}	235 _{II}
	TNT	Wild-type A1(I) A2‡ (I)	K S S	Q R R	E T E	W A A	F A A	E K K	C A S	D A A	D I I	M R R	L Q Q	T S S	R D D	N L L	M A A	N N	N F F	D T T
	Lac	A1(A) A2(A)	K K	K K	L Y Y	Y Y Y	S	S	A A A	S S S	S S S	R R R	R R R	S S S	S S S	S S S	T T T	N K N	K H N	
	Stn	A1(I) A2‡ (I)	A A	Q Q	E E	A A	S S	Q Q	S S	D D	D N	E E	L L	T T	R R	E E	M M	N N	N N	
HBP			11 _I	14 _I	52 _I	70 _I	71 _I	72 _I	77 _I	117 _{II}	120 _{II}	121 _{II}	122 _{II}	143 _{II}	161 _{II}					
	TNT	Wild-type H1(I) H2(I)	D T N	Y S A	L N N	S	L F Q	S K K	R R R	L A L	T S S	T	Q	Q Q I	D S N S I					
	Lac	H1(A) H2(A)	S S	S S	L K	S T	S	K K K	R R R	K H	S S	S A	Q T T	I T	I T					
QBP			10 _I	13 _I	50 _I	70 _I	75 _I	115 _{II}	118 _{II}	156 _{II}	157 _{II}	185 _{II}								
	Lac	Wild-type Q1(A) Q2(A) Q3(A)	D D D	F L L	F L K	T S S	R R D	K K K	T T S	H H H	D H R	Y F F								

The Nomenclature for the designed receptors gives the target ligand and a single-letter abbreviation of the scaffold protein.
 † Subscripts indicate the location of a residue position: I, N-terminal domain; II, C-terminal domain; H, hinge (Fig. 1). Bold letters indicate mutations from wild type (calculations may predict retention of wild-type residues). Underlined letters represent side chains making hydrogen bonds with the ligand (cognate ligand in the case of wild type). The design calculations in a given receptor were not always performed with identical complementary surface residues: automated identification of the complementary surface is indicated by 'A' (see Supplementary Information); 'I' indicates identification by inspection. Blanks therefore indicate residue positions not included in a calculation (wild-type sequence and conformation).
 ‡ TNT.A2 is a point mutant of TNT.A1; Stn.A2 is a point mutant of Stn.A1. Both mutations were designed by inspection.

ligand in a particular scaffold, but also between scaffolds. The likelihood that a protein scaffold can be mutated to accept a new target ligand ('adaptive potential') is also variable. The observed range of affinities can be rationalized with an empirical quantitative structure-activity relationship (QSAR) that provides empirically fitted weights for the DEE force-field components (steric clashes and unsatisfied hydrogen bonds) and takes into account additional factors not modelled by the DEE force field (hydrophobic contact areas, electrostatics and volume ratio of wild-type to target ligands as a measure of adaptive potential; see Supplementary Information). This QSAR provides direct reciprocity between theory and experiment, completing the loop in the protein design cycle, and will be used in future experiments to select scaffolds on the basis of their predicted adaptive potential, and for rank-ordering designs by QSAR-based predicted affinities²².

RBP and GBP control the chemotaxis of *E. coli* towards sugars, which is mediated by a two-component signal transduction pathway²³. This response can be reconnected to gene regulation by constructing a synthetic signal transduction pathway that controls the transcriptional upregulation of a β -galactosidase reporter gene²⁴ (Fig. 3a). The biological activities of the TNT and L-lactate designs in RBP and the L-lactate designs in GBP were tested in this pathway, replacing wild-type RBP and GBP with designed receptors. Wild-type receptors mediate increases in reporter gene expression in response to ribose or glucose but not in response to TNT or L-lactate. Conversely, all the redesigned receptors respond to their target ligands but not to wild-type ligands. The dose-response

Table 2 Affinities of the designed receptors for target ligands and analogues

Target	Receptor	K_d (μ M)			
		TNT	TNB	2,4-DNT	2,6-DNT
TNT	RBP.R1	0.34	1.0	5.0	5.4
	RBP.R2	1.6	3.8	5.3	4.9
	RBP.R3	0.002	0.1	8.4	15
	ABP.A1	1,400	600	>10,000	>10,000
	ABP.A2	400	500	2,000	4,000
	HBP.H1	220	1,000	>10,000	>10,000
		L-Lac		D-Lac	Pyr
L-Lactate	GBP.G1	2.8		205	255
	GBP.G2	2.1		55	115
	HBP.H1	1.8		40	50
	HBP.H2	12.2		30	48
	QBP.Q1	9,500		>100,000	>100,000
	QBP.Q2	300		>100,000	>100,000
	QBP.Q3	25,000		>100,000	>100,000
	ABP.A1	160		>100,000	>100,000
	ABP.A2	20,000		>100,000	>100,000
	RBP.R1	7.4		40	40
		Stn		Trp	Trm
Serotonin	ABP.A1	50		660	900
	ABP.A2	4.7		65	90

The limit of detection for the nitro compounds corresponds to affinities of ~ 10 mM, and for lactate analogues to 100 mM. Error of K_d measurement is approximately 10% (see Supplementary Information). TNT, trinitrobenzene; DNT, dinitrobenzene; Pyr, pyruvate; Stn, serotonin; Trp, L-tryptophan; Trm, tryptamine.

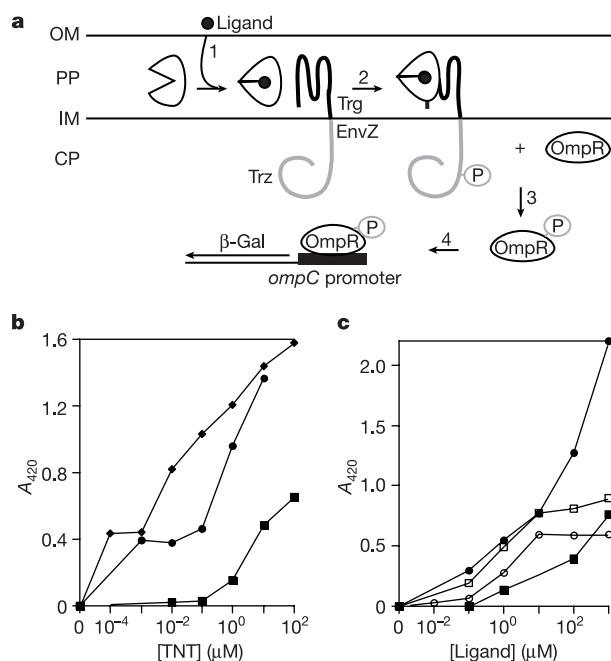


Figure 3 Synthetic two-component signal transduction pathway²⁴. **a**, The ligand-bound RBP or GBP (1) interacts with the Trg domain (thick black line) of a chimaeric transmembrane histidine kinase, Trz (2), resulting in autophosphorylation of the EnvZ domain (grey line) and phosphate transfer to OmpR (3), which then binds to the *ompC* promoter (4), upregulating *lacZ* transcription. OM, outer membrane; PP, periplasm; IM, inner membrane; CP, cytoplasm. **b**, Response to TNT (circles, TNT.R1; squares, TNT.R2; diamonds, TNT.R3). **c**, Response to sugar (open circles, ribose and wild-type RBP; open squares, glucose and wild-type GBP) and L-lactate (filled circles, Lac.R1; filled squares, Lac.G1). β -Galactosidase activities are reported as the difference in assay end-point absorbances of ligand-stimulated and unstimulated cultures. Sensitivity of *E. coli* to high TNT or L-lactate concentrations precluded the determination of full dose-response curves. There was no response in the absence of receptors or *trz*.

curves of the TNT-binding RBP receptors follow the same order as the intrinsic ligand-binding affinities (Fig. 3b, c). The redesigned receptors therefore mediate signal transduction to extracellular TNT or L-lactate, as intended.

We have shown that binding specificities of receptors can be changed drastically by a computational design algorithm. The receptors bind their new cognate ligands selectively relative to closely related decoy molecules, and have affinities that in several cases rival or exceed the affinities of the naturally evolved parent proteins. Essential elements of molecular recognition are therefore correctly captured by the computational design process. The method is relatively rapid and can be applied to many different ligands and proteins. Several of the designed receptors are biologically active, showing that computational design can be used successfully to manipulate biological systems. Furthermore, even within the relatively small number of examples presented here, several potential biotechnological applications are illustrated. The engineered PBPs have the appropriate affinities and specificities for use as reagentless fluorescent biosensors^{5,6} or cell-based sensors²⁵, although further work will be needed to develop devices capable of being deployed in the field. TNT is a potent carcinogen as well as an explosive, and contaminates soil and water near military installations²⁶. The TNT.R3 receptor represents a first step in the development of reagentless fluorescent sensors to detect the source of underwater TNT plumes that leach into the sea from unexploded ordnance in coastal waters, using plume-tracing underwater robotic vehicles²⁷. On land there is an urgent need to detect landmines²⁶. Vapour-phase TNT chemical sensors have been developed successfully²⁸. However, spraying mined fields with bacteria that respond to

TNT by transcriptional activation of a fluorescent reporter gene might offer advantages in cost and simplicity²⁹. The bacterial signal transduction system driven by the engineered TNT.R3 receptor illustrates the potential for developing specific cell-based detectors for chemical threats and pollutants. The L-lactate and serotonin receptors also have practical applications in clinical chemistry. Elevated concentrations of L-lactate are indicative of several medical conditions³⁰. Fluctuations in serotonin blood concentrations are associated with a variety of psychiatric conditions, and elevated concentrations indicate the presence of certain bowel tumours³⁰. The computational design of chirally selective receptors, illustrated by the L-lactate case, could be of use in the facile chromatographic preparation of optically pure pharmaceuticals from racemic mixtures, which is important for drug development⁹. In principle, the computational design strategy presented here can be extended to the design of enzymes^{3,4}. □

Received 4 September 2002; accepted 10 March 2003; doi:10.1038/nature01556.

- Bishop, A. et al. Unnatural ligands for engineered proteins: new tools for chemical genetics. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 577–606 (2000).
- Harris, J. L. & Craik, C. S. Engineering enzyme specificity. *Curr. Opin. Chem. Biol.* **2**, 127–132 (1998).
- Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98**, 14274–14279 (2001).
- Benson, D. E., Haddy, A. E. & Hellinga, H. W. Converting a maltose receptor into a nascent binuclear oxygenase by computational design. *Biochemistry* **41**, 3262–3267 (2002).
- Hellinga, H. W. & Marvin, J. S. Protein engineering and the development of generic biosensors. *Trends Biotechnol.* **16**, 183–189 (1998).
- de Lorimier, R. M. et al. Design, construction and analysis of a family of fluorescent biosensors. *Protein Sci.* **11**, 2655–2675 (2002).
- Hasty, J., McMillen, D. & Collins, J. J. Engineered gene circuits. *Nature* **420**, 224–230 (2002).
- Koh, J. T. Engineering selectivity and discrimination into ligand–receptor interfaces. *Chem. Biol.* **9**, 17–23 (2002).
- Maier, N. M., Franco, P. & Lindner, W. Separation of enantiomers: needs, challenges, perspectives. *J. Chromatogr. A* **906**, 3–33 (2001).
- Arnold, F. H. Combinatorial and computational challenges for biocatalyst design. *Nature* **409**, 253–257 (2001).
- Fersht, A. R. *Structure and Mechanism in Protein Science* (Freeman, New York, 1999).
- Dahiyat, B. I. & Mayo, S. L. *De novo* protein design: Fully automated sequence selection. *Science* **278**, 82–87 (1997).
- Looger, L. L. & Hellinga, H. W. Generalized dead-end elimination algorithms make large-scale protein side-chain structure prediction tractable: implications for protein design and structural genomics. *J. Mol. Biol.* **307**, 429–445 (2001).
- Reina, J. et al. Computer-aided design of a PDZ domain to recognize new target sequences. *Nature Struct. Biol.* **9**, 621–627 (2002).
- Tam, R. & Saier, M. H. Jr. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. *Microbiol. Rev.* **57**, 320–346 (1993).
- Vyas, M. N., Vyas, N. K. & Quiocho, F. A. Crystallographic analysis of the epimeric and anomeric specificity of the periplasmic transport/chemosensory protein receptor for D-glucose and D-galactose. *Biochemistry* **33**, 4762–4768 (1994).
- Mowbray, S. L. & Cole, L. B. 1.7 Å X-ray structure of the periplasmic ribose receptor from *Escherichia coli*. *J. Mol. Biol.* **225**, 155–175 (1992).
- Quiocho, F. A. & Vyas, N. K. Novel stereospecificity of the L-arabinose-binding protein. *Nature* **310**, 381–386 (1984).
- Sun, Y. J., Rose, J., Wang, B. C. & Hsiao, C. D. The structure of glutamine-binding protein complexed with glutamine at 1.94 Å resolution: comparisons with other amino acid binding proteins. *J. Mol. Biol.* **278**, 219–229 (1998).
- Yao, N., Trakhanov, S. & Quiocho, F. A. Refined 1.89 Å structure of the histidine-binding protein complexed with histidine and its relationship with many other active transport/chemosensory proteins. *Biochemistry* **33**, 4769–4779 (1994).
- Kuntz, I. D., Chen, K., Sharp, K. A. & Kollman, P. A. The maximal affinity of ligands. *Proc. Natl Acad. Sci. USA* **96**, 9997–10002 (1999).
- Dahiyat, B. I. & Mayo, S. L. Protein design automation. *Protein Sci.* **5**, 895–903 (1996).
- Stock, A. M., Robinson, V. L. & Goudreau, P. N. Two-component signal transduction. *Annu. Rev. Biochem.* **69**, 183–215 (2000).
- Baumgartner, J. W. et al. Transmembrane signalling by a hybrid receptor: communication from the domain of chemoreceptor Trg that recognizes sugar-binding proteins to the kinase/phosphatase domain of osmosensor. *EnvZ. J. Bacteriol.* **176**, 1157–1163 (1994).
- Daunert, S. et al. Genetically engineered whole-cell sensing systems: coupling biological recognition with reporter genes. *Chem. Rev.* **100**, 2705–2738 (2000).
- Yinon, Y. Field detection and monitoring of explosives. *Trends Anal. Chem.* **21**, 292–301 (2002).
- Mead, K. S. Using lobster noses to inspire robot sensor design. *Trends Biotechnol.* **20**, 276–277 (2002).
- Cumming, C. L. et al. Using novel fluorescent polymers as sensory materials for above-ground sensing of chemical signature compounds emanating from buried land mines. *IEEE Trans. Geosci. Remote Sens.* **39**, 1119–1128 (2001).
- Burlage, R. S., Patek, D. R. & Everman, K. R. Method for detection of buried explosives using a biosensor. US patent 5,972,638 (1999).
- Burtis, C. A. & Ashwood, E. R. *Tietz Textbook of Clinical Chemistry* (Saunders, London, 1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank M. Inouye for the gift of the RU1012 strain, L. Loew for the gift of styryl dyes, S. Conrad and G. Shirman for assistance with mutagenesis and protein chemistry, and M. G. Prisant for construction of the computer cluster. This work was supported by grants from the Office of Naval Research, the Defense Advanced Research Project Agency and the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for material should be addressed to H.W.H. (hwh@biochem.duke.edu).

Direct observation of catch bonds involving cell-adhesion molecules

Bryan T. Marshall*, Mian Long*†, James W. Piper*†, Tadayuki Yago‡, Rodger P. McEver‡§ & Cheng Zhu*||

* Woodruff School of Mechanical Engineering and || Coulter School of Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA
‡ Cardiovascular Biology Research Program, Oklahoma Medical Research Foundation, and § Department of Biochemistry and Molecular Biology and Oklahoma Center for Medical Glycobiology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma 73104, USA

Bonds between adhesion molecules are often mechanically stressed. A striking example is the tensile force applied to selectin–ligand bonds, which mediate the tethering and rolling of flowing leukocytes on vascular surfaces^{1–3}. It has been suggested that force could either shorten bond lifetimes, because work done by the force could lower the energy barrier between the bound and free states⁴ ('slip'), or prolong bond lifetimes by deforming the molecules such that they lock more tightly^{5,6} ('catch'). Whereas slip bonds have been widely observed^{7–14}, catch bonds have not been demonstrated experimentally. Here, using atomic force microscopy and flow-chamber experiments, we show that increasing force first prolonged and then shortened the lifetimes of P-selectin complexes with P-selectin glycoprotein ligand-1, revealing both catch and slip bond behaviour. Transitions between catch and slip bonds might explain why leukocyte rolling on selectins first increases and then decreases as wall shear stress increases^{9,15,16}. This dual response to force provides a mechanism for regulating cell adhesion under conditions of variable mechanical stress.

Using atomic force microscopy (AFM) (Fig. 1a), we measured the force dependence of bond lifetimes of P-selectin with two forms of P-selectin glycoprotein ligand-1 (PSGL-1) or with G1, a blocking monoclonal antibody (mAb) against P-selectin¹⁷ (see Methods). P-selectin is an extended C-type lectin expressed on activated endothelial cells and platelets. PSGL-1 is a mucin expressed on leukocytes. Ca²⁺-dependent interactions of P-selectin with PSGL-1 mediate the tethering and rolling of flowing leukocytes on vascular surfaces in response to infection or tissue injury^{1–3}.

We captured dimeric PSGL-1 purified from human neutrophils¹⁸ or monomeric recombinant soluble PSGL-1 (sPSGL-1)¹⁹ with PL2, a non-blocking anti-PSGL-1 mAb²⁰ adsorbed on the cantilever tip (Fig. 1b). Cantilever tips bearing (s)PSGL-1 or G1 were repeatedly brought into contact with lipid bilayers reconstituted with P-selectin purified from human platelets²¹ to allow bond formation. The cantilever was then retracted a prescribed distance to apply a constant tensile force to the bond or bonds (if any resulted from the contact), and the duration or lifetime of the adhesion at that

force was recorded (Fig. 1c). To measure lifetime at forces lower than the level of their fluctuations, many instantaneous forces were averaged (Fig. 1d, e). This enabled the reliable resolution of mean forces as low as a few piconewtons, and allowed the detected differences in mean forces to achieve high statistical significance (Fig. 1f). The binding frequency was kept low (12–20%) to ensure that most (about 90%) adhesions dissociated as a single step (Fig. 1c, lower tracing). Only single-step dissociations were analysed.

Binding was highly specific. Rendering the cantilever tip functional with (s)PSGL-1 increased adhesion frequencies 3–10-fold (Fig. 2a and b) and also increased bond lifetimes (Fig. 3a and b). The inclusion of blocking mAbs against P-selectin (G1) or against PSGL-1 (PL1²⁰) or the divalent-cation chelator EDTA in the chamber solution decreased adhesion to nonspecific levels. G1-coated cantilever tips had significantly higher adhesion frequencies and longer bond lifetimes than control PL2-coated tips (Figs 2c and 3c).

Remarkably, both the P-selectin–sPSGL-1 interaction (Fig. 3a) and the P-selectin–PSGL-1 interaction (Fig. 3b) exhibited a biphasic relationship between lifetime and force. The bond lifetimes initially increased with force, indicating the presence of catch bonds. After reaching a maximum, the lifetimes decreased with force, indicating slip bonds. This biphasic pattern was detected in individual experiments with a single cantilever tip on a single bilayer, which included as few as about five lifetime measurements to calculate a mean and a standard deviation at each of four force levels to cover the biphasic region. This pattern remained unchanged as more data were accumulated (about 400 lifetime measurements for each form of PSGL-1), which allowed us to examine the distributions of lifetimes. As with published flow-chamber data^{7,9,11–13}, lifetimes at a given force seemed to follow an exponential distribution, which was made linear by plotting $\ln(\text{number of events with a lifetime of } t \text{ or more})$

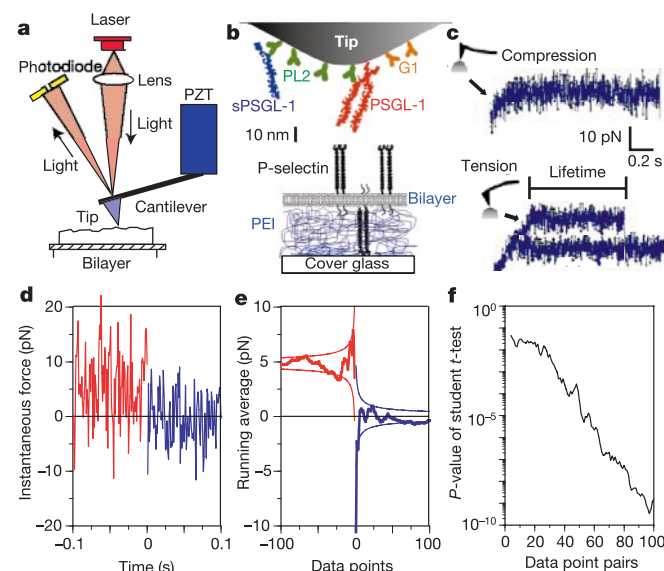


Figure 1 AFM system. **a**, Schematic diagram. **b**, Making the AFM functional. The cantilever tip depicted represents a composite of all molecules adsorbed or captured. sPSGL-1 and PSGL-1 are depicted as monomer and dimer, respectively. **c**, Force-scan curves illustrating the cantilever bending (insets) when a compressive or tensile force was applied to the tip. The upper curves illustrate a contact cycle without binding, where the retraction curve (from left to right) retraced the approach curve (from right to left). The lower curves illustrate a contact cycle with binding and lifetime measurement. **d**, Plot of instantaneous force against time before (red) and after (blue) an unbinding event. **e**, Running means (thick curves) $\pm$ s.e.m. (paired thin curves) against number of data points. **f**, *P* value of the Student's *t*-test comparing the difference of the two running means against numbers of point pairs.

† Present addresses: National Microgravity Laboratory, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China (M.L.), and Immucor, Inc., Norcross, Georgia 30091, USA (J.W.P.).

against t ; examples are given in Fig. 3d, e. The negative reciprocal of the slope of the linear fit was used as a lifetime measure and was plotted against force with the 95% confidence interval as error bar. For first-order dissociation kinetics, $-1/\text{slope}$ equals the mean, $\langle t \rangle$, and standard deviation, $\sigma(t)$, of lifetimes, which were also plotted against force. The steepness of plots of $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t decreased with increasing force in the ranges that corresponded to the ascending phase of the curves of $\langle t \rangle$ against f and of $\sigma(t)$ against f for P-selectin interacting with sPSGL-1 (Fig. 3d) and with PSGL-1 (Fig. 3e). The differences between any two neighbouring points in these ranges (where most of the data reside) were statistically significant ($P < 0.04$), as determined by the F -test comparing the slopes of the two plots of $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t , and confirmed by the lack of overlaps in the confidence intervals. Thus, the P-selectin-(s)PSGL-1 bonds behaved as catch bonds in the low force range. To our knowledge, this is the first experimental demonstration of a catch bond. Our data also indicate that a molecular interaction can change from a catch bond in one force range to a slip bond in another force range—a phenomenon that existing models have not predicted.

We compared the force dependence of lifetimes of the P-selectin-(s)PSGL-1 bonds with that of the P-selectin-G1 bond in the same force range. Other assays have revealed significant differences between selectin–ligand interactions and antigen–antibody interactions that are subjected to tensile force²². In sharp contrast to the P-selectin–PSGL-1 bond lifetimes (Fig. 3a, b), the P-selectin–G1 bond lifetimes decreased precipitously with force, displaying typical slip bond characteristics that are well described by the Bell model⁴ (Fig. 3c). These qualitatively distinct relationships demonstrate that the catch–slip transitional bond is specific for the P-selectin–PSGL-1 interaction.

The curves of lifetime against force for the two forms of PSGL-1 had similar biphasic shapes, but the PSGL-1 curve (Fig. 3b) was shifted relative to the sPSGL-1 curve (Fig. 3a), approximately doubling the forces and lifetimes. As depicted in Figs 1b and 4a, native membrane P-selectin and PSGL-1 are dimeric, whereas sPSGL-1 is monomeric^{12,19,23}. The AFM data suggest that sPSGL-1 forms monomeric bonds with P-selectin, whereas PSGL-1 forms dimeric bonds with P-selectin, which was further supported by molecular elasticity measurements (B.T.M., K. K. Sarangapani, A. Leppänen, R. D. Cummings, R.P.McE. and C.Z., unpublished observations). A dimeric bond would allow the applied force to be spread between two subunits, thereby supporting larger forces and

shifting the curve rightwards. In addition, the cantilever tip would not detach from the bilayer until both subunits of the dimeric bond had dissociated, thereby prolonging the lifetime and shifting the curve upwards. Indeed, the lifetime data in Fig. 3a, b can be fitted by the same force dependence of off-rate, if the sPSGL-1 data are analysed by a model of single-step dissociation of monomeric bonds and the PSGL-1 data are analysed by a model of two-step dissociation of dimeric bonds (B.T.M., V. Ramachandran, R.P.McE. and C.Z., unpublished observations). These combined results strongly suggest that the AFM measured single monomeric P-selectin–sPSGL-1 bonds and single dimeric P-selectin–PSGL-1 bonds. The ability to differentiate between single monomeric and dimeric bonds attests to the sensitivity of the lifetime assay and to the reliability of the catch–slip transitional bond results. Furthermore, it indicates that the measurements reflect lifetimes of the P-selectin-(s)PSGL-1 bonds, not of the (s)PSGL-1–PL2 bonds or the P-selectin–bilayer anchors.

With the use of flow chambers, several groups have shown that lifetimes of tethers of PSGL-1-expressing cells to P-selectin-coated surfaces decreased with increasing wall shear stress, which is consistent with slip bond behaviour^{7,9,11–13}. These experiments were conducted at wall shear stresses ranging from 0.17 to 3.0 dyn cm⁻², corresponding to tether forces of 20–400 pN. In addition, AFM was previously used to study P-selectin–PSGL-1 interactions by loading the bonds with increasing forces until rupture, which suggested a slip bond at forces of more than 50 pN (ref. 24). Our AFM data revealed that the P-selectin–sPSGL-1 and P-selectin–PSGL-1 interactions behaved as catch bonds only at

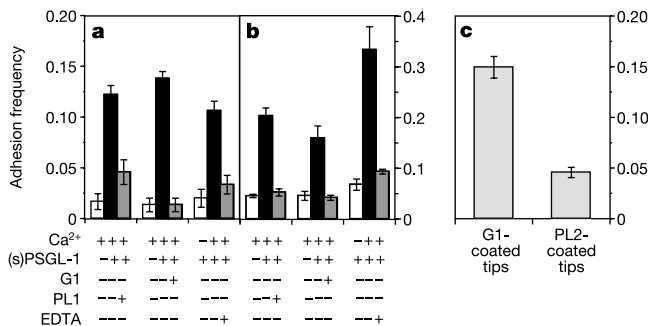


Figure 2 Binding specificity. **a, b**, The P-selectin bilayer was tested by a cantilever tip with sequentially changing treatments: contacts were first made with no (s)PSGL-1 on the tip or no Ca²⁺ in the medium (open bars); the same tip was then made functional with sPSGL-1 (**a**) or PSGL-1 (**b**), or the medium was changed to include Ca²⁺ (solid bars); finally, EDTA or the indicated mAb was added to the medium (hatched bars). **c**, The P-selectin bilayer was contacted with a tip adsorbed with either G1 or PL2. Adhesion frequency for each condition was measured from three to five spots with 50 contacts each on a single bilayer–tip pair and results are presented as means $\pm$ s.e.m. for at least three separate bilayer–tip pairs.

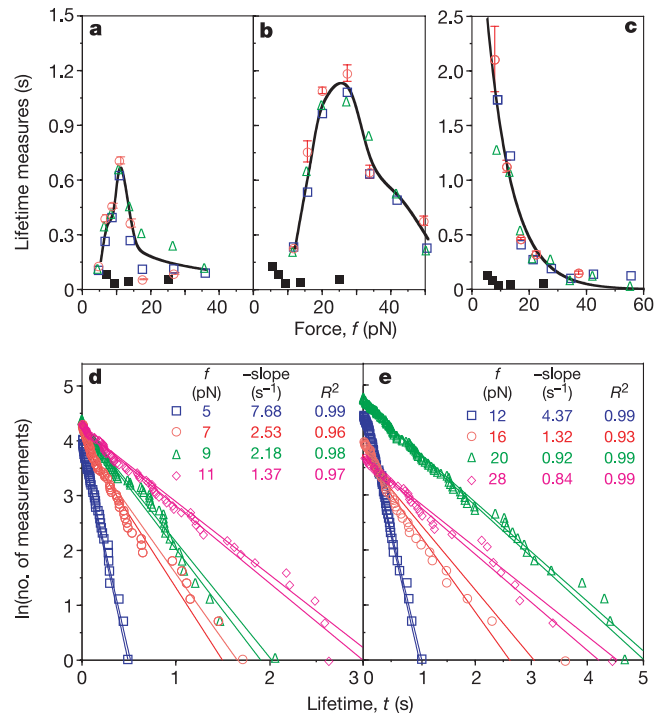


Figure 3 Lifetimes measured by AFM. **a–c**, Statistical measures of nonspecific mean lifetimes (black filled squares), mean lifetimes $\langle t \rangle$ (open blue squares), standard deviation of lifetimes $\sigma(t)$ (green triangles) and $-1/\text{slope}$ of the plot of $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t (red circles), for bonds of P-selectin with sPSGL-1 (**a**), PSGL-1 (**b**) or G1 (**c**). The nonspecific lifetimes were measured with tips adsorbed with PL2 only. The curves are trend lines in **a** and **b**, but in **c** it is a Bell model fit, $\langle t \rangle = \langle t \rangle_0 \exp(-af/k_B T)$, where $\langle t \rangle_0 = 4.6$ s, $a = 0.46$ nm, k_B is Boltzmann constant, and T is absolute temperature⁴. **d, e**, Plots of $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t in the catch bond regime for sPSGL-1 (**d**) and PSGL-1 (**e**). The 95% confidence intervals of the slopes are shown by the paired, colour-matched lines, also shown as error bars in **a–c**.

forces below 10 and 20 pN, respectively. To address whether catch bonds could be observed in the flow chamber, we perfused neutrophils or microspheres coupled with sPSGL-1 or G1 at low wall shear stresses over sparsely coated P-selectin, which is dimeric, or over recombinant soluble P-selectin (sP-selectin), which is monomeric²³ (Fig. 4a). Measurements of transient tether lifetimes confirmed catch bonds for (s)P-selectin–(s)PSGL-1 interactions in the force range coinciding with that of the AFM experiments. Both sPSGL-1-coupled microspheres and neutrophils displayed a biphasic force–lifetime relationship (Fig. 4b, left and middle panels). The curves of $-1/\text{slope}$ against f were identical for sPSGL-1-coated microspheres tethering to P-selectin or sP-selectin (Fig. 4b, left panel), as predicted from the inability of sPSGL-1 to form dimeric bonds with P-selectin. In contrast, the curve of $-1/\text{slope}$ against f for neutrophils tethering to P-selectin was shifted rightwards and upwards relative to that for neutrophils tethering to sP-selectin, approximately doubling the forces and lifetimes (Fig. 4b, middle panel). These corroborate the AFM data (Fig. 3a, b) and confirm that neutrophil PSGL-1 formed dimeric bonds with P-selectin but not with sP-selectin¹². The specific nature of the (s)P-selectin–(s)PSGL-1 interactions was confirmed by contrasting their behaviour with that of the G1-coupled microspheres, which exhibited a force dependence of lifetimes that was consistent only with slip bonds (Fig. 4b, right panel). The observation of catch–slip transitional bonds in both AFM and flow-chamber studies strongly suggests that this counterintuitive behaviour is not an experimental artefact.

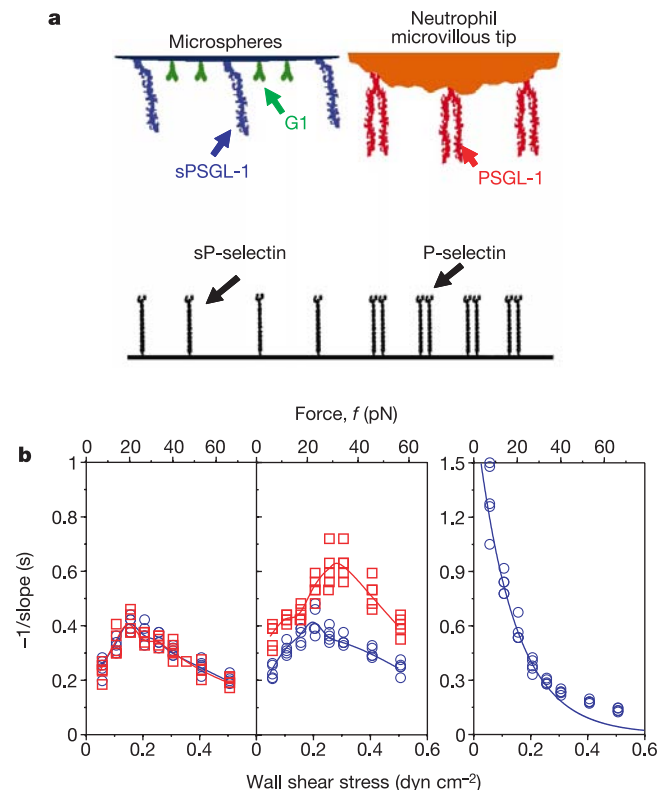


Figure 4 Lifetimes measured in a flow chamber. **a**, Schematic diagram showing microspheres coupled with either sPSGL-1 or G1, neutrophils expressing PSGL-1 on microvillous tips, and dimeric P-selectin or monomeric sP-selectin adsorbed on plastic. **b**, Five sets of data of $-1/\text{slope}$ of $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t plots were plotted against wall shear stress (bottom abscissa) and force on a tether (top abscissa) for sPSGL-1 microspheres (left panel), neutrophils (middle panel) and G1 microspheres (right panel) tethering to P-selectin (red squares) or sP-selectin (blue circles). The colour-matched curves are trend lines through the mean data in the left and middle panels; the right panel shows a Bell model⁴ fit, with best-fit parameters $\langle t \rangle_0 = 1.8 \text{ s}$ and $a = 0.24 \text{ nm}$.

P-selectin binds in a stereospecific manner to an amino-terminal region of PSGL-1 through recognition of tyrosine sulphate residues, adjacent peptide determinants, and fucose and sialic acid residues on a properly positioned core-2 O-glycan^{3,25}. The crystal structure of this complex reveals a broad, shallow binding interface, but it does not readily explain how force elicits the catch–slip transitional bond features that we observed²⁶.

The use of both catch and slip bonds might be physiologically important. Leukocytes must experience a minimum threshold wall shear stress to tether to and roll on selectins^{9,15,16}. As wall shear stress is increased, the number of rolling cells first increases and then decreases in a biphasic pattern similar to our lifetime data. The wall shear stresses studied were higher than those that evoked catch–slip transitional bond behaviour. However, multiple bonds might mediate most leukocyte tethers, which could reduce the force on individual bonds to the levels that produce the catch–slip bond transition^{10,12}. Bonds formed at the leading edge of the rolling cell are gradually loaded by force as they move to the trailing edge. These bonds will therefore experience forces in the catch bond regime even at high shear stresses at the wall. At low wall shear stresses, bond lifetimes would be too brief to support rolling. When a shear threshold is reached, catch bonds would retard dissociation at the trailing edge, thereby stabilizing rolling. Catch bonds might explain why platelet tethering to von Willebrand factor on disrupted vessels is most effective at higher shear stresses at the arterial wall³. Catch bonds have also been proposed to regulate flow-dependent bacterial adhesion²⁷. Transitions between catch and slip bonds might provide a general mechanism for the precise regulation of cell adhesion during mechanical stress. □

Methods

Lifetimes measured by AFM

Our custom-made AFM (design adapted from V. Moy, University of Miami) used commercial cantilevers (TM Microscopes, Sunnyvale, CA) with spring constants from 4 to 13 pN nm⁻¹, calibrated by the thermal fluctuation method²⁸. PSGL-1 or sPSGL-1 was captured by anti-PSGL-1 mAb PL2 adsorbed on a tip, which was then blocked in Hanks buffer with 1% BSA. Anti-P-selectin mAb G1 was adsorbed on a separate tip followed by blocking. P-selectin was reconstituted in a polyethylenimine (PEI) polymer-supported lipid bilayer^{29,30}. The PEI cushion reduced nonspecific adhesion by accommodating the inversely inserted P-selectin. Low molecular densities were used to achieve infrequent binding. Binding was enabled by using a piezoelectric translator (PZT) to bring the tip to contact the bilayer for 3 s with a ~ 20 -pN compressive force. Force signals were collected from the photodiode, which measured the laser reflected on the back of the cantilever. The cantilever was then retracted at 250 nm s⁻¹ and stopped once it had arrived at a predetermined distance to apply a constant force to the bond or bonds. The lifetime was measured from the instant when the PZT stopped to the instant of failure of the bond or bonds.

Force averaging was employed to allow lifetime measurement at forces comparable to the level of thermal fluctuations (4–10 pN), as illustrated in Fig. 1d–f. The instantaneous force immediately before unbinding could not be used as the force at which the lifetime was measured because it was highly variable (see Fig. 1d for an example). However, the running mean $\langle f \rangle$ stabilized after a few tens of fluctuating forces were included, and $\langle f \rangle \approx 4.86 \pm 0.52 \text{ pN}$ was calculated from 100 points for a lifetime of 0.1 s (data collected at a rate of 1 kHz). Another 100 post-rupture points yielded $\langle f \rangle \approx 0.00 \pm 0.42 \text{ pN}$. In both cases, the standard deviations remained at $\sim 5 \text{ pN}$ but the s.e.m. decreased with increasing data points (Fig. 1e). This enabled a reliable resolution of the two mean forces and allowed the detected differences to achieve high statistical significance, as indicated by the rapidly decreasing P value of Student's t -test comparing the difference of the two running means with increasing numbers of point pairs included in the calculation (Fig. 1f).

Lifetimes measured at forces ranging from 5 to 75 pN were divided into six to eight force bins and analysed by three statistical measures: mean lifetime $\langle t \rangle$, standard deviation of lifetime $\sigma(t)$ and $-1/\text{slope}$ of the plot of $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t . They are related to kinetic parameters by means of models for the dissociation of a monomeric bond,

$$B_1 \xrightarrow{k_{-1}} R_1 + L_1$$

or of a dimeric bond,

$$B_2 \xrightleftharpoons[k_{-2}]{k_{+2}} R_1^* + L_1^* + B_1 \xrightarrow{k_{-1}} R_2 + L_2$$

where R , L and B denote receptor, ligand and bond, respectively, k_+ and k_- denote on-rates and off-rates, respectively, the superscript $*$ indicates the unbound but constrained leg of a dimeric molecule, and subscripts 1 and 2 denote monomeric and dimeric molecules or processes, respectively. For lifetimes measured at constant force f , k_{-1} was

assumed to depend on force but not time. $k_{-2}(f) = 2k_{-1}(f/2)$ because the force on a dimeric bond is shared equally by the two subunits, each of which behaves as a monomeric bond and is equally likely to dissociate. k_{+2} is assumed to be a constant. The monomeric model predicts that lifetimes are distributed exponentially. Performing a log-linear plot renders the exponential distribution linear; thus $-k_{-1}$ equals the slope of the $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t plot. Also, $\langle t \rangle_1 = \sigma_1(t) = 1/k_{-1}$. By comparison, the dimeric model predicts that

$$\langle t \rangle_2 = (1 + k_{+2}/k_{-2})/k_{-1} + 1/k_{-2}$$

and

$$\sigma_2(t) = \{[(1 + k_{+2}/k_{-2})/k_{-1} + 1/k_{-2}]^2 - 2/(k_{-1}k_{-2})\}^{1/2}$$

The $k_{-1}(f)$ values that allow the monomeric model to fit the data in Fig. 3a also allow the dimeric model to fit the data in Fig. 3b.

Lifetimes measured in a flow chamber

Lifetimes of transient tethers of neutrophils and microspheres were measured in flow-chamber experiments as described in refs 12 and 19. Neutrophils were isolated from healthy donors. Polystyrene microspheres (6 μm diameter; Polysciences, Warrington, Pennsylvania) either precoated or not with streptavidin were incubated with biotinylated PSGL-1 or mAb G1. Neutrophils or microspheres were perfused at various wall shear stresses over dimeric P-selectin or monomeric sP-selectin adsorbed at low densities on the chamber floor. The tethers were eliminated by coating the floor with human serum albumin instead of (s)P-selectin or by the inclusion of the anti-P-selectin mAb G1, the anti-PSGL-1 mAb PL1 or EDTA in the medium, confirming their specificity. The force f on the tether of a neutrophil or a microsphere was calculated on the basis of the tether angle, which was derived after directly measuring the lever arm of the tether by a flow reversal method¹⁹. The conversion factors between wall shear stress and tether force for neutrophils and microspheres are 112 and 131 $\text{pN dyn}^{-1} \text{cm}^{-2}$, respectively. Five sets of lifetimes (~ 100 tether events in each set) were measured for each interaction at each wall shear stress. Each set was analysed by the plot of $\ln(\text{number of events with a lifetime of } t \text{ or more})$ against t , which was fitted by a straight line. The correlation coefficients, R^2 , were more than 0.9 for all 200 fits. The $-1/\text{slope}$ values of the fits were plotted against the wall shear stress (points in Fig. 4b). The difference in mean $-1/\text{slope}$ values at any two neighbouring wall shear stresses was statistically significant ($P < 0.05$, Student's t -test), except occasionally at the beginning of the slip bond regime after transition from the catch bond regime.

Received 11 November 2002; accepted 19 February 2003; doi:10.1038/nature01605.

- Vestweber, D. & Blanks, J. E. Mechanisms that regulate the function of the selectins and their ligands. *Physiol. Rev.* **79**, 181–213 (1999).
- McEver, R. P. & Cummings, R. D. Perspectives series: cell adhesion in vascular biology. Role of PSGL-1 binding to selectins in leukocyte recruitment. *J. Clin. Invest.* **100**, 485–491 (1997).
- McEver, R. P. Adhesive interactions of leukocytes, platelets, and the vessel wall during hemostasis and inflammation. *Thromb. Haemost.* **86**, 746–756 (2001).
- Bell, G. I. Models for the specific adhesion of cells to cells. *Science* **200**, 618–627 (1978).
- Dembo, M., Tournay, D. C., Saxman, K. & Hammer, D. The reaction-limited kinetics of membrane-to-surface adhesion and detachment. *Proc. R. Soc. Lond. B* **234**, 55–83 (1988).
- Dembo, M. *Lectures on Mathematics in the Life Sciences, Some Mathematical Problems in Biology* 51–77 (American Mathematical Society, Providence, Rhode Island, 1994).
- Alon, R., Hammer, D. A. & Springer, T. A. Lifetime of the P-selectin–carbohydrate bond and its response to tensile force in hydrodynamic flow. *Nature* **374**, 539–542 (1995).
- Pierres, A., Benoliel, A. M., Bongrand, P. & van der Merwe, P. A. Determination of the lifetime and force dependence of interactions of single bonds between surface-attached CD2 and CD48 adhesion molecules. *Proc. Natl Acad. Sci. USA* **93**, 15114–15118 (1996).
- Alon, R., Chen, S., Puri, K. D., Finger, E. B. & Springer, T. A. The kinetics of L-selectin tethers and the mechanics of selectin-mediated rolling. *J. Cell Biol.* **138**, 1169–1180 (1997).
- Chen, S. & Springer, T. A. An automatic braking system that stabilizes leukocyte rolling by an increase in selectin bond number with shear. *J. Cell Biol.* **144**, 185–200 (1999).
- Smith, M. J., Berg, E. L. & Lawrence, M. B. A direct comparison of selectin-mediated transient, adhesive events using high temporal resolution. *Biophys. J.* **77**, 3371–3383 (1999).
- Ramachandran, V. et al. Dimerization of a selectin and its ligand stabilizes cell rolling and enhances tether strength in shear flow. *Proc. Natl Acad. Sci. USA* **98**, 10166–10171 (2001).
- Chen, S. & Springer, T. A. Selectin receptor–ligand bonds: Formation limited by shear rate and dissociation governed by the Bell model. *Proc. Natl Acad. Sci. USA* **98**, 950–955 (2001).
- Merkel, R., Nassoy, P., Leung, A., Ritchie, K. & Evans, E. Energy landscapes of receptor–ligand bonds explored with dynamic force spectroscopy. *Nature* **397**, 50–53 (1999).
- Finger, E. B. et al. Adhesion through L-selectin requires a threshold hydrodynamic shear. *Nature* **379**, 266–269 (1996).
- Lawrence, M. B., Kansas, G. S., Kunkel, E. J. & Ley, K. Threshold levels of fluid shear promote leukocyte adhesion through selectins (CD62L, P-E). *J. Cell Biol.* **136**, 717–727 (1997).
- Geng, J.-G. et al. Rapid neutrophil adhesion to activated endothelium mediated by GMP-140. *Nature* **343**, 757–760 (1990).
- Moore, K. L. et al. The P-selectin glycoprotein ligand from human neutrophils displays sialylated, fucosylated, O-linked poly-N-acetylglucosamine. *J. Biol. Chem.* **269**, 23318–23327 (1994).
- Yago, T. et al. Distinct molecular and cellular contributions to stabilizing selectin-mediated rolling under flow. *J. Cell Biol.* **158**, 787–799 (2002).
- Moore, K. L. et al. P-selectin glycoprotein ligand-1 mediates rolling of human neutrophils on P-selectin. *J. Cell Biol.* **128**, 661–671 (1995).
- Moore, K. L., Varki, A. & McEver, R. P. GMP-140 binds to a glycoprotein receptor on human neutrophils: evidence for a lectin-like interaction. *J. Cell Biol.* **112**, 491–499 (1991).
- Chen, S., Alon, R., Fuhbrigge, R. C. & Springer, T. A. Rolling and transient tethering of leukocytes on antibodies reveal specializations of selectins. *Proc. Natl Acad. Sci. USA* **94**, 3172–3177 (1997).
- Ushiyama, S., Laue, T. M., Moore, K. L., Erickson, H. P. & McEver, R. P. Structural and functional

characterization of monomeric soluble P-selectin and comparison with membrane P-selectin. *J. Biol. Chem.* **268**, 15229–15237 (1993).

- Fritz, J., Katopodis, A. G., Kolbinger, F. & Anselmetti, D. Force-mediated kinetics of single P-selectin/ligand complexes observed by atomic force microscopy. *Proc. Natl Acad. Sci. USA* **95**, 12283–12288 (1998).
- Leppänen, A., White, S. P., Helin, J., McEver, R. P. & Cummings, R. D. Binding of glycosulfopeptides to P-selectin requires stereospecific contributions of individual tyrosine sulfate and sugar residues. *J. Biol. Chem.* **275**, 39569–39578 (2000).
- Somers, W. S., Tang, J., Shaw, G. D. & Camphausen, R. T. Insights into the molecular basis of leukocyte tethering and rolling revealed by structures of P- and E-selectin bound to SLex(X) and PSGL-1. *Cell* **103**, 467–479 (2000).
- Isberg, R. R. & Barnes, P. Dancing with the host; flow-dependent bacterial adhesion. *Cell* **110**, 1–4 (2002).
- Hutter, J. L. & Bechhoefer, J. Calibration of atomic-force microscope tips. *Rev. Sci. Instrum.* **64**, 1868–1873 (1993).
- McConnell, H. M., Watts, T. H., Weis, R. M. & Brian, A. A. Supported planar membranes in studies of cell-cell recognition in the immune system. *Biochim. Biophys. Acta* **864**, 95–106 (1986).
- Wong, J. Y. et al. Polymer-cushioned bilayers. I. A structural study of various preparation methods using neutron reflectometry. *Biophys. J.* **77**, 1445–1457 (1999).

Acknowledgements We thank V. Moy for providing the AFM design and training. This work was supported by grants from the National Institutes of Health.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.Z. (cheng.zhu@me.gatech.edu).

Folding at the speed limit

Wei Yuan Yang* & Martin Gruebele*†

* Center for Biophysics and Computational Biology, and † Departments of Chemistry and Physics, University of Illinois, Urbana Illinois 61801, USA

Many small proteins seem to fold by a simple process explicable by conventional chemical kinetics and transition-state theory. This assumes an instant equilibrium between reactants and a high-energy activated state¹. In reality, equilibration occurs on timescales dependent on the molecules involved, below which such analyses break down¹. The molecular timescale, normally too short to be seen in experiments, can be of a significant length for proteins. To probe it directly, we studied very rapidly folding mutants of the five-helix bundle protein λ_{6-85} , whose activated state is significantly populated during folding. A time-dependent rate coefficient below 2 μs signals the onset of the molecular timescale, and hence the ultimate speed limit for folding². A simple model shows that the molecular timescale represents the natural pre-factor for transition state models of folding.

The pre-factor $\nu^\ddagger$, together with the activation energy $\Delta G^\ddagger$, determines the rate coefficient $k = \nu^\ddagger \exp(-\Delta G^\ddagger/kT)$ of transition state theory. For small molecules, $\nu^\ddagger \approx kT/h \approx (0.2 \text{ ps})^{-1}$ is a reasonable value. Proteins are simply too large to move about and fold in a fraction of a picosecond. Protein folding pre-factors have been estimated on the basis of random intramolecular collision within small peptide loops, yielding $\nu^\ddagger \approx 10\text{--}100 \text{ ns}$ (refs 3–6). This number might be too small for the molecular timescale of larger proteins. Their backbone and side chains can collide in more ways and with more non-additive interactions, introducing additional roughness into the free-energy landscape. A calculation for the five-helix bundle protein λ_{6-85} has yielded numbers closer to 0.5 μs (refs 7, 8), and evidence from unfolded cytochrome c in denaturant indicates values in the 1–40- μs range⁹. Single-molecule experiments have put an upper limit of 200 μs on the pre-factor¹⁰.

This range of pre-factors leaves us with an ambiguity that has so far precluded the determination of protein folding barriers from experiment: many combinations of $\nu^\ddagger$ and $\Delta G^\ddagger$ yield the same

observable rate coefficient k . Without sufficient knowledge of the critical reaction coordinates for describing the motions represented by $\nu^\ddagger$, it is impossible to relate experimentally determined folding rates rigorously to computed free-energy barriers.

We provide direct evidence that the molecular timescale of λ_{6-85} is 2 μ s, by observing relaxation towards equilibrium induced by a fast temperature jump induced by a laser¹¹. Rigorous kinetic theory has shown that the rate coefficient increases from its phenomenological value once the molecular timescale is reached¹. Below about 4 μ s, such a speed-up beyond exponential kinetics occurs for our fastest-folding mutants. Slower-folding proteins have barriers that are too high to reveal the depletion of activated populations, and show perfect single-exponential kinetics. We analyse our results in terms of numerical diffusion calculations with a simple one-dimensional two-well model, whose renormalized diffusion coefficient accounts for traps and reduced dimensionality. (Models with explicit traps also fit the data; see Supplementary Information and ref. 12.)

The wild type and many mutants of the 80-residue amino-terminal λ repressor fragment λ_{6-85} (including sites Asp 14, Gln 33, Ala 37, Gly 46 and Gly 48 (refs 13–15), and T. Oas, personal communication) have been studied by Oas and co-workers. They

introduced the Tyr22Trp mutation as a fluorescence probe that shifts spectrally from 334 to 355 nm after the unfolding and solvent exposure of the protein core¹⁴. An NMR line shape analysis¹³ and stopped-flow measurements¹⁴ in denaturant solutions both revealed rates that extrapolated to less than 20 μ s for the fastest mutants in the absence of any denaturant.

Such fast folding makes λ_{6-85} an ideal starting point for creating even faster mutants to reach the speed limit. Unlike NMR studies, our temperature-jump experiments can be performed with nanosecond resolution under very favourable folding conditions without denaturant. All our proteins contained the mutations Tyr22Trp and Gln33Tyr. The Tyr 33/Trp 22 pair introduces an aromatic interaction that speeds up folding and provides a large Trp fluorescence-lifetime change after unfolding. The three distinguishing sets of mutations were: Gly46Ala/Gly48Ala, which speeds up folding by stabilizing helix 4; Asp14Ala/Gly46Ala/Gly48Ala, an additional mutation that removes a hydrogen bond between positions 14 and 77; and Ala37Gly, a model for slower folding because helix 4 is not stabilized and helix 3 is destabilized instead. We abbreviate these mutants below as λ Q33Y, λ D14A and λ A37G (see Methods).

The three variants had helical contents very similar to the wild type, judging from their circular dichroism spectra (Fig. 1a). Thermal denaturation occurs by means of sigmoidal transitions of the circular dichroism and fluorescence spectra between 1 and 95 °C, confirming apparent two-state thermodynamic behaviour (Fig. 1a inset). The unfolded states are exposed to solvent and have largely lost helical structure. Denaturant titration (Fig. 1b) reveals at most a small displacement between different spectroscopic probes.

In the kinetics experiments, laser temperature jumps from 9–20 °C to a series of final temperatures were applied to initiate relaxation (see Methods). Relaxation to the new equilibrium constant was monitored by tryptophan fluorescence lifetimes. The excellent signal-to-noise ratio allowed us to measure relaxation rates even under folding conditions, with final equilibrium constants $K > 20$. The kinetics of λ A37G was the slowest ($k_a \approx 50 \mu$ s), and fitted well to a single exponential under all experimental conditions (Fig. 2c). The faster-folding λ Q33Y and λ D14A are single-exponential only at $t > 4 \mu$ s. Their relaxation transients speed up beyond the main relaxation phase below 4 μ s. This very fast phase grows to more than 20% of the total signal under the most favourable folding conditions (lowest temperatures, Fig. 2a, b). λ Q33Y relaxation is fairly well fitted by a double exponential (Figs 2a and 3a), indicating that there is still some separation of timescales in the underlying dynamics. The even faster λ D14A data are better fitted by a stretched exponential $\exp(-kt)^\beta$, with $\beta = 0.72 \pm 0.02$ (Fig. 2b), indicating a merger of the two timescales as folding becomes faster. For easy comparison, Fig. 3a reports double-exponential fits of k_m (molecular rate coefficient) and k_a (activated rate coefficient) for all proteins, and Fig. 3b shows activated folding rate coefficients k_f .

The origin of the swift initial decay is of particular interest because the transient lies in the bracket between 10 ns and 40 μ s set for the molecular timescale by peptide-diffusion^{3,4}, unfolded-protein^{9,16} and single-molecule experiments¹⁰. The initial decay speeds up the kinetics when optimal folding conditions are approached, as predicted by rigorous rate theory¹. It is an 'anti-rollover' of the folding rate, not a slow-down caused by a rate-limiting folding intermediate. The data cannot be globally fitted to a model involving a conventional early intermediate (see Supplementary section 2). The ratio of the fast to the main phase does not increase when the temperature-jump size is increased from 9 to 20 °C. λ A37G has no fast phase, so equilibration in the folded and unfolded well, which should occur in slow proteins also, is ruled out as the sole source of the signal. The fast mutants have no fast phase when jumped from one native condition to another, ruling out equilibration within the folded well as the only cause. For the fastest mutants, the main folding phase is slowed down at high concen-

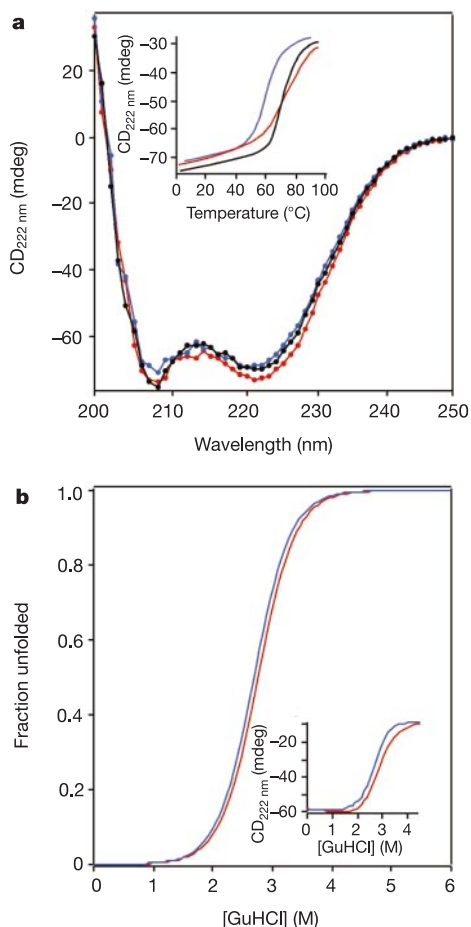


Figure 1 Spectroscopy of mutant proteins. **a**, Circular dichroism (CD) signatures of the three λ_{6-85} mutants (20 °C, 5 μ M, 1-cm cuvette): λ Q33Y (red), λ D14A (black) and λ A37G (blue). Inset: sigmoidal temperature titration curves. **b**, Unfolded population fits on guanidinium chloride (GuHCl) titration data (assuming that the free energy is linearly dependent on guanidinium chloride concentrations) taken by circular dichroism at 222 nm (red) and fluorescence at 280 nm (blue). Inset: raw data. Baseline-corrected population fits show at most small deviations of the transition midpoint by different measurement techniques.

trations by aggregation¹⁷ (Fig. 2d). We therefore operate in the low-concentration limit ($[\lambda A37G] < 300 \mu M$, $[\lambda Q33Y] < 100 \mu M$ and $[\lambda D14A] < 40 \mu M$), where the main folding phase has converged to its fastest value, in accord with a previous model¹⁷. The lower concentration thresholds of the fastest mutants indicate partly unfolded populations. This could be because partly folded preactivated conformations are already populated in the lowest barrier mutants, and interact in quantity upon collision.

Having excluded other reasonable explanations, we assign the fast phase to the increase in speed beyond phenomenological kinetics predicted by theory below the molecular timescale¹. A one-dimensional two-well model can account for all the observed kinetic and thermodynamic trends of all three mutants as a function of temperature (Fig. 4; see Methods and Supplementary Information). In this model, a low barrier supports both a preactivated subpopulation that reacts on the very fast molecular timescale, as well as a main population that reacts on the activated timescale. Higher barriers (slower-folding proteins) yield only the activated single-exponential relaxation signal of conventional two-state models.

Simulations were performed by equilibrating populations with the use of Langevin dynamics in the high friction limit, then applying a jump to the potential and monitoring population re-equilibration (see Methods and Supplementary Information). To reproduce the molecular and activated phases and their relative amplitudes, the effective diffusion constant must be decreased to $1.6 \times 10^{-4} \text{ nm}^2 \text{ ns}^{-1}$, and the barrier is very low for the fast mutants (Fig. 4).

Because we can observe both the molecular-timescale k_m^{-1} and activated kinetics in one measurement, we can devise a useful equation for the activation energy. The molecular rate constant k_m in Fig. 3 reflects the average time that proteins spend in the activated region and can therefore be related to the slower activated rate constant k_a by

$$k_a/k_m = \exp(-G_{\text{act}}/kT) \quad (1)$$

This equation is an operational definition of barrier size in terms of the energetically accessible phase space in the activated as opposed to the well regions of the free energy. Thus, a very broad higher barrier can have the same population as a lower narrower

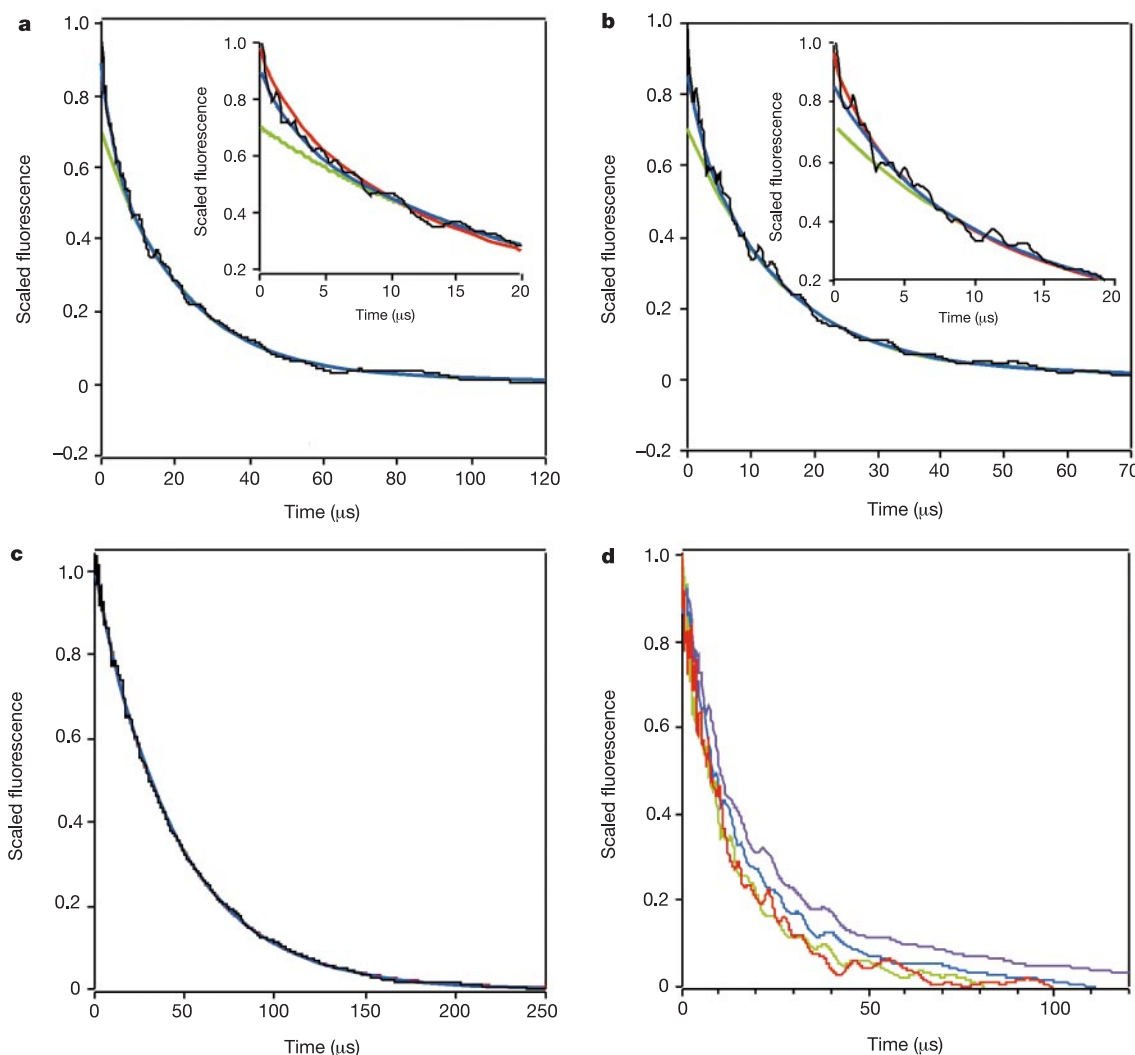


Figure 2 Relaxation after temperature jumps. **a**, Relaxation of $\lambda Q33Y$ at 63 °C (black curve). Blue curve, double exponent; green curve, long time exponent. Inset: detailed view near $t = 0$. A stretched exponential (red; $\beta = 0.74 \pm 0.02$) performs comparably to a double-exponential fit (black). The best overall exponential fit (blue) and long-time exponential fit (green, slow component of double exponential fit) are also shown, and deviate markedly from experiment. **b**, As in **a**, for $\lambda D14A$ at 61 °C. Here the stretched-exponential fit ($\beta = 0.72 \pm 0.02$) performs slightly better than the double-exponential

fit. **c**, The slower $\lambda A37G$ mutant at 56 °C (black curve) fits to a single exponential (blue curve), even when the stretching factor β is allowed to vary ($\beta = 0.96 \pm 0.04$). The fast mutants both require $\beta < 0.8$. **d**, Transient aggregation slows the relaxation at high concentrations of $\lambda D14A$, at 61 °C. Below 40 μM (for $\lambda D14A$) and 100 μM (for $\lambda Q33Y$; see Supplementary Information), the relaxation converges to a maximum rate reflecting the collision-free folding reaction. Red curve, 15 μM ; green curve, 31 μM ; blue curve, 95 μM ; purple curve, 173 μM .

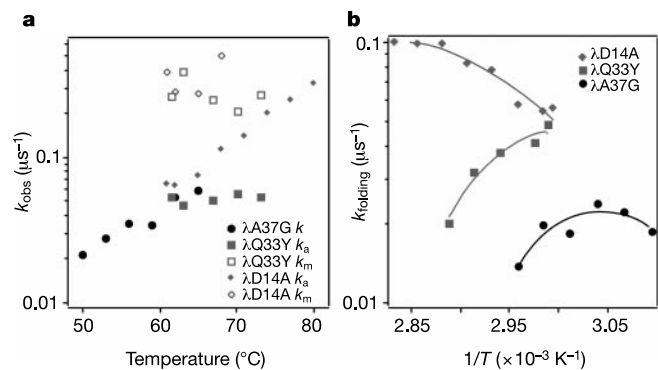


Figure 3 Rate coefficient determination. **a**, Single-exponential and double-exponential fits to the experimental relaxation data. Two rate coefficients (k_m , molecular; k_a , activated) are shown for the fast mutants, and one (k_a) for the slow mutant. The separation of timescales ranges from only 4 to 8 for λD14A and λQ33Y . **b**, Activated folding rates computed from k_a and the known equilibrium constants K (Fig. 1) from $k_f = k_a K / (K + 1)$. Quadratic activation energy fits are shown. The fast mutants behave very differently from one another, with the slow mutant showing intermediate turnover behaviour (see Supplementary Information).

barrier. Our definition, based on the separation of experimental timescales, yields $G_{\text{act}} \approx \ln(8.3)kT \approx 2.1kT$ for λQ33Y at 60°C , and $G_{\text{act}} \approx \ln(4.5)kT \approx 1.5kT$ for λD14A at 65°C . The value $1.5kT$ lies outside the range of validity of activated rate models. Previously, only partial downhill folding and thermodynamic signs for barrier-free folding were observed^{18,19}.

Why is the $k_m^{-1} \approx 2 \mu\text{s}$ molecular timescale slower than the measured 50-ns loop closure time for unfolded peptides? When an 80-residue protein folds, it undergoes multiple 50-ns contact formations and rearrangements along the reaction coordinate: it has a broad, rough activated region. Broad-barrier models, and also dimpled-barrier models—which lump all the roughness into one state (see Supplementary Information)—have been proposed^{20,21}. A more realistic picture is that the barrier region contains a distribution of local minima, the deepest of which can be classified as intermediates. Such a picture has been found^{7,8}, with multiple mini-transition states from an analytical folding model for λ_{6-85} . A pre-factor for backbone dynamics of about $0.5 \mu\text{s}$ has been calculated for the wild-type protein under favourable folding conditions, in good agreement with our measured molecular timescale^{7,8}. Our renormalized diffusion constant of about $1.6 \times 10^{-4} \text{ nm}^2 \text{ ns}^{-1}$ (compared with about $0.05 \text{ nm}^2 \text{ ns}^{-1}$ in inter-chain diffusion measurements⁹) mimics an even distribution of traps and landscape roughness, and corrects for the reduced dimensionality of the model. A $4.7kT$ trap and $0.007 \text{ nm}^2 \text{ ns}^{-1}$ diffusion constant must be invoked to account for the roughness in a dimpled-barrier model (see Supplementary Information).

The speed limit for λ_{6-85} is $2 \mu\text{s}$. Whether this is ascribed to a renormalized diffusion constant, a single trap or multiple local minima along an explicitly computed reaction coordinate depends on the level of theory chosen. A barrier size given by the ratio of the molecular to the activated timescale (equation (1)) is a more robust descriptor for protein folding than the barrier height and curvature usually chosen for small molecules. Fortunately, analytical theory⁷, and now a direct comparison of experimental folding kinetics and molecular dynamics simulations²², avoid over-reliance on phenomenological rate models. However, simulations must exceed the molecular timescale to determine rate coefficients reliably. For very small proteins, a nanosecond timescale is probably sufficient²², but larger proteins evidently require about $1 \mu\text{s}$ or longer before phenomenological rate equations can be applied. Finally, fast folding phases of some smaller proteins, interpreted more conventionally in terms of unfolded states under native conditions or

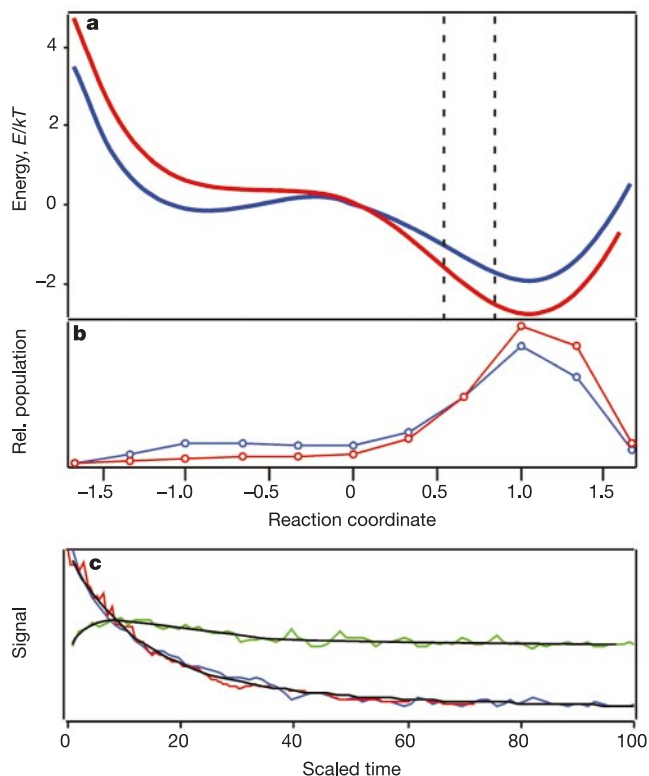


Figure 4 One-dimensional free-energy model used to reproduce the λQ33Y and λD14A data as a function of temperature. **a**, Double-well surface before (red) and after (blue) temperature jump to 63°C ; vertical bins mark where the fluorescence signal switches from unfolded to native in the model. **b**, Binned equilibrated protein populations before and after the jump, showing sizeable activated populations. **c**, Population (green line) and double-exponential fit (black curve) for the bin in **a**, and λQ33Y experimental kinetics (red line), simulated fluorescence signal (blue line) and double-exponential fit (black curve). The relative rates and amplitudes fit λD14A at 63°C (and at all other temperatures (not shown); see Supplementary Information); the absolute rates are matched by adjusting the diffusion constant in the Langevin equation (see the text).

folding intermediates²³, need careful further investigation in view of the results and models outlined here. □

Methods

Experimental

The gene encoding λ_{6-85} was a gift from T. Oas. It was inserted into the PET-15b vector, tagging the protein with six histidine residues followed by a thrombin cleavage site at the N terminus. Mutations were carried out using the Quickchange site-directed mutagenesis kit (Stratagene). Proteins were expressed in BL21(DE3) and purified through a Ni^{2+} -nitrilotriacetate and a Sephacryl S-200 HR column (Pharmacia). Thrombin was used to remove the hexahistidine tag. The final λ -repressor used in the measurements contained three extra residues (Gly-Ala-Met) at the N terminus of the wild-type sequence. All measurements were done in 50 mM potassium phosphate buffer, pH 7. Concentrations were determined assuming an extinction coefficient of $8,200 \text{ cm}^{-1} \text{ M}^{-1}$ at 280 nm. Circular dichroism spectra were taken at $5 \mu\text{M}$ protein concentration, fluorescence spectra at $1 \mu\text{M}$. The melting points of the mutants were 60°C for λA37G , 71°C for λQ33Y and 74°C for λD14A . Kinetics were monitored by exciting tryptophan at 288 nm every 14 ns with a mode-locked, frequency-tripled Ti:S laser. The total fluorescence was filtered through a 320–380-nm bandpass filter, imaged onto a 500-ps rise-time photomultiplier tube, and digitized every 0.5 ns. Scaled fluorescence was calculated by fitting the fluorescence decay shape as $f(t) = \chi_1 f_1 + (1 - \chi_2) f_2$, where f_1 is the initial fluorescence profile just after the temperature jump and f_2 is the fluorescence profile after equilibration. This yields a single exponential only when a two-state model applies; otherwise it deviates from single-exponential behaviour²⁴. Magnitudes of phases can be compared quantitatively as a function of temperature, but not between mutants because of innate differences in the fluorescence profiles. We have engineered a low-stability variant containing mutations Ala37Gly/Ala49Gly in addition to Tyr22Trp that can be denatured by cold. It can be both refolded (from the cold-denatured state at subzero temperatures) and unfolded (from the folded state at high temperatures) with folding rates differing only by the ratio of the water viscosity at the two temperatures. Because water viscosity has a

large inherent dependence on temperature, we concluded that the folding reaction is at least near the Kramers high friction limit.

Computational

The double-well potential used was $g(x) = x^4 - 2x^2$. Temperature effects were treated as a linear bias along the reaction coordinate, $t(x, T) = A(T)x$, where $A(T)$ is an adjustable parameter for matching the equilibrium data. The folded and unfolded states are separated by 2 distance units along the reaction coordinate, corresponding to a typical helix diffusion length when taken to be nanometres. The population at $x < 0.83$ was assumed to have the same fluorescence signature as the unfolded state, and at $x > 0.83$ as the folded state (for compatibility with the three-well model in the Supplementary Information, any value $x > 0$ yields the same qualitative result). Fluorescence was simulated by convolving populations (for example Fig. 4b) with this response. The one-dimensional Langevin equation with gaussian white noise was integrated by using a fourth-order Runge-Kutta method. A time-step size of 0.01 was used in the integration, and time steps were scaled to match the experimentally observed absolute kinetics. Similar calculations for a three-well model that also matches the data are described in the Supplementary Information.

Received 24 January; accepted 31 March 2003; doi:10.1038/nature01609.

- Berne, B. J. in *Activated Barrier Crossing: Applications in Physics, Chemistry and Biology* (eds Hänggi, P. & Fleming, G. R.) 82–119 (World Scientific, Singapore, 1993).
- Hagen, S. J., Hofrichter, J., Szabo, A. & Eaton, W. A. Diffusion-limited contact formation in unfolded cytochrome c: Estimating the maximum rate of protein folding. *Proc. Natl Acad. Sci. USA* **93**, 11615–11617 (1996).
- Lapidus, L. J., Eaton, W. A. & Hofrichter, J. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl Acad. Sci. USA* **97**, 7220–7225 (2000).
- Bieri, O. *et al.* The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl Acad. Sci. USA* **96**, 9597–9601 (1999).
- Yeh, I. C. & Hummer, G. Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J. Am. Chem. Soc.* **124**, 6563–6568 (2002).
- Thirumalai, D. Time scales for the formation of the most probable tertiary contacts in proteins with applications to cytochrome c. *J. Phys. Chem. B* **103**, 608–610 (1999).
- Portman, J. J., Takada, S. & Wolynes, P. G. Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics. *J. Chem. Phys.* **114**, 5082–5096 (2001).
- Portman, J. J., Takada, S. & Wolynes, P. G. Microscopic theory of protein folding rates. I. Fine structure of the free energy profile and folding routes from a variational approach. *J. Chem. Phys.* **114**, 5069–5081 (2001).
- Hagen, S. J., Hofrichter, J. & Eaton, W. A. Rate of intrachain diffusion of unfolded cytochrome c. *J. Phys. Chem. B* **101**, 2352–2365 (1997).
- Schuler, B., Lipman, E. A. & Eaton, W. A. Probing the free-energy surface for protein folding with single-molecule fluorescence spectroscopy. *Nature* **419**, 743–747 (2002).
- Gruebele, M., Sabelko, J., Ballew, R. & Ervin, J. Laser temperature jump induced protein refolding. *Acc. Chem. Res.* **31**, 699–707 (1998).
- Metzler, R., Klafter, J. & Jortner, J. Hierarchies and logarithmic oscillations in the temporal relaxation patterns of proteins and other complex systems. *Proc. Natl Acad. Sci. USA* **96**, 11085–11089 (1999).
- Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L. & Oas, T. G. The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions. *Nature Struct. Biol.* **4**, 305–309 (1997).
- Ghaemmaghami, S., Word, J. M., Burton, R. E., Richardson, J. S. & Oas, T. G. Folding kinetics of a fluorescent variant of monomeric lambda repressor. *Biochemistry* **37**, 9179–9185 (1998).
- Myers, J. M. & Oas, T. G. Contribution of a buried hydrogen bond to lambda repressor folding kinetics. *Biochemistry* **38**, 6761–6768 (1999).
- Chang, I. J., Lee, J. C., Winkler, J. R. & Gray, H. B. The protein-folding speed limit: intrachain diffusion times set by electron transfer rates in denatured Ru(NH₃)₅(His-33)-Zn-cytochrome c. *Proc. Natl Acad. Sci. USA* **100**, 3838–3840 (2003).
- Silow, M. & Oliveberg, M. Transient aggregates in protein folding are easily mistaken for folding intermediates. *Proc. Natl Acad. Sci. USA* **94**, 6084–6086 (1997).
- Sabelko, J., Ervin, J. & Gruebele, M. Observation of strange kinetics in protein folding. *Proc. Natl Acad. Sci. USA* **96**, 6031–6036 (1999).
- Garcia-Mira, M. M., Sadqi, M., Fischer, N., Sanchez-Ruiz, J. M. & Muñoz, V. Experimental identification of downhill protein folding. *Science* **298**, 2191–2195 (2002).
- Silow, M. & Oliveberg, M. High-energy channeling in protein folding. *Biochemistry* **36**, 7633–7637 (1997).
- Pappenberger, G., Saudan, C., Becker, M., Merbach, A. E. & Kiefhaber, T. Denaturant-induced movement of the transition state of protein folding revealed by high-pressure stopped-flow measurements. *Proc. Natl Acad. Sci. USA* **97**, 17–22 (2002).
- Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. Absolute comparison of simulated and experimental protein-folding dynamics. *Nature* **420**, 102–106 (2002).
- Mayor, U. *et al.* The complete folding pathway of a protein from nanoseconds to microseconds. *Nature* **421**, 863–867 (2003).
- Ervin, J., Sabelko, J. & Gruebele, M. Submicrosecond real-time fluorescence detection: application to protein folding. *J. Photochem. Photobiol. B* **54**, 1–15 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Oas for suggesting many helpful lambda repressor mutations. This work was supported by an Alumni Scholarship and a Camille Dreyfus Teacher-Scholar Award to M.G.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G. (gruebele@scs.uiuc.edu).

erratum

Antibody neutralization and escape by HIV-1

Xiping Wei, Julie M. Decker, Shuyi Wang, Huxiong Hui, John C. Kappes, Xiaoyun Wu, Jesus F. Salazar-Gonzalez, Maria G. Salazar, J. Michael Kilby, Michael S. Saag, Natalia L. Komarova, Martin A. Nowak, Beatrice H. Hahn, Peter D. Kwong & George M. Shaw

Nature **422**, 307–312 (2003).

In the seventh panel of Fig. 2 of this Letter, the V5 sequence of clone 391-3 appeared incorrectly as: SEKDQTEIFRP. It should read: SKDNQTEIFRP. In addition, there should be no yellow shading (indicating a change in glycosylation) for this sequence. □

corrigenda

Synaptic depression in the localization of sound

Daniel L. Cook, Peter C. Schwindt, Lucinda A. Grande & William J. Spain

Nature **421**, 66–70 (2003).

It has come to our attention that we failed to cite a relevant study¹ in our Letter. These authors identified the mechanism of synaptic depression measured at the embryonic chick nucleus magno-cellularis to nucleus laminaris synapse as primarily presynaptic, which justifies the synaptic depletion model we used. Furthermore, the narrowing of coincidence detection time windows with EPSP depression as they observed may contribute to the adaptive mechanisms that we described. □

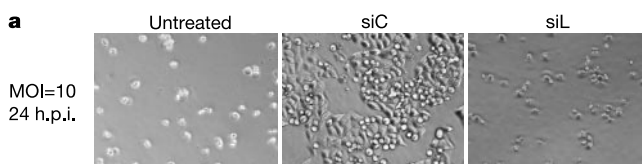
- Kuba, H., Konomi, K. & Ohmori, H. *Eur. J. Neurosci.* **15**, 984–990 (2002).

Short interfering RNA confers intracellular antiviral immunity in human cells

Leonid Gitlin, Sveta Karelsky & Raul Andino

Nature **418**, 430–434 (2002).

In Fig. 5a of this Letter, the first and third panels (untreated and siL-treated cells, respectively) should not be identical: the correct figure is shown here. □



Microbiology miscellany

Bugs, cells, counts and sample preparation.



Look no glass: MantaRay, from Wheaton.

MantaRay

Wheaton Science Products

www.wheatonsci.com

Throw-away culture

The MantaRay pre-sterilized, disposable spinner flask is designed as an inexpensive alternative to glass spinner flasks. It consists of a free-standing plastic container with a fill cap and sealed ports for pumps or injectors. The unit has a working volume range of 0.4–1 litre and includes an integral stirring mechanism. The user simply fills the flask with media, adds cells, grows the culture and disposes of the entire unit on completion of the experiment. The MantaRay can be used with a wide range of animal, plant and insect cells for suspension culture or adherent cells on microcarrier beads.

EasySep

StemCell Technologies www.stemcell.com

Positive attraction

EasySep is a column-free magnetic system that can select CD34⁺ cells in about an hour. Cells are positively selected by incubating a single cell suspension (cord blood, bone marrow, mobilized peripheral blood) with the Easy Sep CD34 selection cocktail and magnetic nanoparticles in a standard FACS tube. The tube is placed into a hand-held magnet and unwanted cells are poured off, while the selected CD34⁺ cells remain in the tube. The magnetic nanoparticles do not interfere with subsequent FACS analysis and therefore do not need to be removed.

Bactozol

Molecular Research Center

www.mrcgene.com

Bacterial DNA isolation kit

Bactozol can isolate DNA from a broad spectrum of Gram-negative and Gram-positive bacteria. The kit features Bactozyme, a concentrated solution containing activated lysozyme that remains stable at room temperature and precludes the need for daily enzyme preparation, and DNAzol, which provides a fast and efficient method for isolation of DNA from the bacterial lysate. The kit includes sufficient reagents for 125 isolation procedures, each yielding up to 40 µg of DNA.

BIO 101 Transform & Grow

Qbiogene

www.qbiogene.com

E. coli preparation with ease

Qbiogene has introduced a bacterial transformation kit for the preparation, storage and transformation of chemically competent *E. coli* host cells. The kit comprises Transform-It solution, liquid Super-Comp media, and Roll & Grow plating beads. It is designed to provide subcloning and plasmid transfer with a minimum of hands-on time and without the need for the heat shock procedure often associated with 'do-it-yourself' cell preparation and storage. Qbiogene recommends the kit for use with its BIO 101 bacterial growth media and GeneClean-based plasmid purification kits.

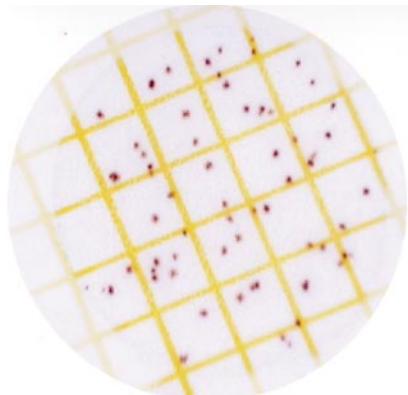
Petrifilm plate holder

Synbiosis

www.synbiosis.com

Support for automated ProtoCOL

Synbiosis manufactures a plate holder for use with its ProtoCOL automated colony



Quick diagnosis: counting on Petrifilm.

counter. The Petrifilm plate holder ensures correct magnification of any type of colony grown on Petrifilm and performs an automatic colony count. It is easily integrated into the ProtoCOL system and, according to Synbiosis, will save time and improve accuracy when performing tests for harmful bacteria such as coliforms, Enterobacteriaceae and *Staphylococcus aureus*.

iCyte and LSC cytometers

Olympus Optical www.olympus-europa.com

Laser-sharp imaging tools

Olympus has added CompuCyte's iCyte imaging cytometer and LSC laser scanning cytometer to its range of products for biomedical research. The iCyte is an automated, multi-laser instrument for high-content cell and tissue analysis suitable for automated screening processes in biomedicine and drug discovery applications. The iCyte features automated cellular analysis of multiple specimens simultaneously in a variety of formats, including microtitre plates and microscope slides. The LSC laser scanning cytometer is an automated, high data-content cell analyser for specimens on microscope

slides. Both instruments provide precise measurements of morphological and multiple quantitative cell features.

MACS microscope

Don Whitley Scientific www.dwscientific.co.uk

Held up for inspection

This stereo microscope from Don Whitley Scientific can be used on any modular atmosphere controlled systems (MAC) workstation. The microscope attachment allows a magnified area of a Petri dish, hand-held for inspection inside the cabinet, to be viewed three-dimensionally. When not in use, the microscope can be left in an almost 'ready-to-use' status, yet stored out of the way so as not to interfere with normal workstation operations.

Stuart SC 6

Bibby Sterilin www.bibby-sterilin.com

Colony counter with built-in averaging

The Stuart SC colony counter has a digital readout of 0 to 999 and can count multiple plates, calculating the average colony count via a built-in average count facility. The



Cell-based assays: the Olympus/CompuCyte combination.



Stuart SC6: 37 and counting.

device features an adjustable, pressure-sensitive count system that responds to any probe or felt-tip pen, providing an audible confirmation of each count. It also offers a choice of light or dark background, with peripheral, glare-free illumination. It comes with a Wolfhuegel graticule, a segmentation disc and a centering adapter for 50–90 mm dishes. A magnifier is available as an optional extra.

Mortar grinder RM100 and KM100

Retsch www.retsch-us.com

Cool and collected

Retsch claims that yeast cell extraction is improved by the cryogenic grinding process used in the company's RM100 and KM100 mortar grinders. The addition of liquid nitrogen during the grinding maintains a constant sample temperature, ensuring a smooth extraction without RNA damage. The RM100 grinder handles up to 150 ml and the KM 100 up to 300 ml of sample.

Ruskinn Concept 1000

Biotrace International www.biotrace.com

Two's company

Designed for high-throughput and research microbiology laboratories, the Concept 1000 incubation system features a central interlock that serves both chambers independently, allowing multi-user access and control of different temperature and humidity environments as required. Users can work side by side using different conditions within the dual gas environment. The fully automated interlock transfers plates at a rate of 100 per 5 minutes, with an economy cycle for smaller loads. The unit has a total capacity of 1,000 plates, 500 in each work chamber. Ezee Sleeve ports with bare-hand access facilitate working with samples under anaerobic conditions, and a 15 cm single-plate entry system provides instant access to each work chamber.

Online showcase

Alexis Corporation www.alexis-e.biz

Check out the goods in the window

This one-stop information and purchasing platform provides access to the reagents and services of original biochemical products manufacturers. It lists more than 10,000 biochemicals, immunochemicals and assay kits in several major categories, including apoptosis, cell cycle, cytoskeletal research, immunology, cancer research, neurochemicals, nitric oxide and signal transduction. Visitors to the site can quickly locate specific life science reagents and view all the products available in an area of research. Data sheets, manuals, catalogues and information can be viewed, downloaded or printed. All of the products can be ordered on-line, and technical and customer support is also available.

Flowmaster FMT 300

Ismatec/Micropump www.micropump.com

Be gentle with it

The Ismatec Flowmaster peristaltic pump features a convex roller and concave tube-bed designed to transfer media accurately and gently. The tubing is compressed progressively from the centre outwards, causing a temporary gap that permits the cells to escape before being squeezed. Cell viability remains high and cell damage low, ensuring a high concentration of viable cells for inoculation. The single-channel pump has a maximum flow rate of 13 litres per minute. It has IP65 protection for safe dispensing and filling in a dusty, humid or corrosive environment, and in clean room areas.

These notes are compiled in the Nature office from information provided by the manufacturers. Material intended for publication can be sent to newproducts@nature.com.



Joining the Ruskinn Bugbox in the range, this is the Concept 1000 incubator.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/

Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel +49 89 54 90 57 11/-2

Fax +49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatomachi,
Shinjuku-ku,
Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Cash crisis in the corridor

Michigan's dreams of carving out a niche for itself in biotechnology have suffered a temporary set-back. Financed with settlement money obtained from a multi-state lawsuit against the tobacco industry, Michigan's Life Sciences Corridor had been building itself an attractive reputation (see *Naturejobs* 4-5; 21 November 2002). But after three years of investing \$50 million a year, Michigan is having to make cutbacks in the face of mounting economic problems.

What once looked like a relatively stable long-term source of funding has tightened. Because the state budget was in trouble, Michigan's governor, Jennifer Granholm, felt the need to cut costs. She reduced the corridor's funding to \$32 million for 2003, and has proposed reducing it to \$20 million in 2004 (see *Nature* 422, 102; 2003).

Two facets of the programme will emerge relatively unscathed, however. Michigan will continue with its five-year plan to invest in core technologies, such as a nuclear-magnetic-resonance facility, and will also continue to spend \$5 million a year on seed money for start-ups. But the biggest component, grants to investigators, is being gutted, although the programme's director hopes that the annual funding of \$50 million will be restored when the state's economy recovers.

The situation in Michigan is worth bringing to the attention of scientists in other parts of the world, even those who have never considered moving to Michigan. Over the past few years, more and more US states and European countries have started to invest in their regions, promoting them as viable places to do science, in the hope of attracting talent and investment from abroad. Michigan's reversal of fortune, unfortunately, may give them food for thought.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

A fresh outlook for
eastern Germany

p204

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Science has flourished in eastern Germany since reunification, bringing researchers from far and near. But the tide of investment may be turning. Marieke Degen reports.

When Josef Käs set off for Harvard University in 1993, he was one of hundreds of hopeful young German researchers who cross the Atlantic each year searching for a scientific career. Ten years later, he is one of a small group of top scientists for whom the German government is prepared to dig deep in its pockets to entice home.

The letter that made him a member of this élite circle arrived two years ago. In 2001, the government-financed Alexander von Humboldt Foundation offered him the prestigious Wolfgang Paul Award — a scheme designed to bring scientists with international reputations to Germany. The recipients can conduct research in a German lab of their choice, with the unusually generous support of about E2 million (US\$2.2 million) over three years.

Käs didn't think twice. Top universities would have gladly opened their doors to the 40-year-old biophysicist. But the Munich-born researcher made an unexpected move — he established his group at the University of Leipzig in the east of the country.

His decision reflects the growing attractiveness of the country's new eastern *Länder* (states) to young researchers seeking opportunities and first-class working conditions in many emerging fields. Yet Germany's fragile economy risks reversing the east's upswing of the past decade.

PAINFUL TRANSITION

Following German reunification in 1990, the Soviet-style science system of the former East Germany was redesigned into a modern science base (see *Nature* **401**, 635–639; 1999). The transformation, although painful to the thousands of East German researchers who lost their jobs, was largely successful, as was the whole *Aufbau Ost*, a heavily subsidized programme aimed at equalizing economic and living standards in eastern and western Germany.

But not everything has been levelled. Although living costs are lower in the east than in the west, so too are wages, and unemployment is worryingly higher. Also, the number of researchers per thousand in the eastern workforce (3.8) lags behind that in the west (6.0) — mainly because, according to an analysis by the German science ministry, the east still lacks the right

mixture of research-intensive industries of all sizes.

But eastern Germany should no longer fear comparison with the west of the country on the strength of its academic system. The Max Planck Society, Germany's main non-university organization for basic research, for example, has set-up 20 high-profile institutes in the new *Länder*. Cities such as Leipzig, Halle, Jena, Dresden and Greifswald have become attractive science areas, with a good mix of university research, Max Planck institutes, other non-university institutes and science parks.

WELL-EQUIPPED

The recent investments mean that — contrary to general belief — scientists in eastern Germany often have more modern instrumentation and tools in their labs than their counterparts in the west.

Eric Marois, a French microbiologist at the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, says that instrumentation was an attraction, not an obstacle, in his decision to move east. “We have everything you could desire,” says Marois. “Our microscopes, centrifuges, animal rooms and anaesthetic stations are absolutely top-notch.”

Technical and animal-care staff are plentiful and competent, he says, which he finds particularly helpful. Also, foreign researchers are assisted by a special employee who deals with bureaucratic formalities, which can be a tedious duty, particularly for those whose first language is not German.

There is some xenophobia, occasionally reported by the media, but this isn't a reason to keep away from Germany, says Marois. “There have been some incidents of violence to foreigners, but by and large the risk of being attacked on the street here is certainly no higher than in Paris or in any other big city.”

Marois had the choice four years ago between doing his PhD in the United States or in eastern Germany. He chose Germany's ‘wild east’, which, although geographically nearer to France, culturally and scientifically seemed more distant and adventurous.

“I liked the idea of learning a new language and mentality,” he says. He found the small-town charm of Halle, and later the quiet grandeur of Dresden, more favourable to creative research than the constant rush of Western or Asian capitals.

Josef Käs: enjoys scientific openness, cooperation and a well-equipped lab.





The University of Leipzig's high-energy nanometre ion probe is one of only two in the world.

Käs agrees. He says he likes the open atmosphere of Leipzig, as well as the courage and staying power of its citizens, who precipitated the fall of the Berlin wall in autumn 1989. Now thousands of peace activists in Leipzig have revived the famous Monday demonstrations, this time against the Iraq war.

Another pleasant relic from East Germany, where it was taken for granted that women could work as well as men, is the well-functioning childcare system, which is very convenient for young researchers such as Käs who bring their family.

"I would have come here with or without the Wolfgang Paul prize," he says. "The United States is a great place to live for a while, but I really prefer that my little daughter grows up in old Europe."

There are also many scientific reasons to opt for eastern Germany, he adds. At the university's department of experimental physics, Käs is developing, among other things, a method for the early diagnosis of cancer, by testing the elasticity of cells with laser microscopes. The project involves clinical immunologists from other departments. This level of interaction with colleagues from neighbouring fields is typical in the east, says Käs. At Harvard and the University of Texas at Austin, he seldom encountered cooperation across discipline boundaries.

"If everything has gone well for decades, researchers have little need or incentive to leave their well-known territory and try something new," Käs comments. Now he is experiencing a completely different scientific

spirit. "The institute structures are not yet fixed, and we are basically forced to work closely together," he says. "The borders here are all open, which leads to very productive interdisciplinary research."

STORMCLOUDS ON THE HORIZON

But the academic idyll is under threat. Germany's economic crisis has led to cuts in research budgets that are particularly painful in the east, where massive investment over the past ten years requires high running and repair costs (see *Nature* **420**, 452; 2002).

The physics and Earth-science departments at the University of Leipzig, for example, have expensive nuclear-magnetic-resonance facilities, and a E3.5-million high-energy nanometre ion probe, one of only two of its kind in the world.

"We have almost no money left to keep our excellent equipment running," says Matthias Schaefer, the department's administrative director. The German government has promised to keep up special support of research in the east through programmes such as InnoRegio, which puts E65 million a year into regional academic-industrial research networks.

But the new *Länder* remain the poor cousins when it comes to large research infrastructure. In February, the federal research minister, Edelgard Bulmahn, announced a E1.6-billion investment package into four large projects, including a high-magnetic-field laboratory at the Rossendorf research centre near Dresden, the former East German centre for nuclear research (see *Nature* **421**, 682; 2003). Researchers in the east feel they got a raw deal with the E24.5-million facility. "Almost 98% of the money remains in the west," says Frank Pobell, scientific director of the Rossendorf centre. "If you are serious about the *Aufbau Ost*, the east would have deserved better."

Pobell strongly regrets Germany's vote against the E1.4-billion European Spallation Source (ESS), a proposed major neutron-science facility for which Saxony had already submitted a detailed scientific outline and a site (see *Nature* **421**, 563; 2003). "It is a unique opportunity that has been missed," he says. "The ESS would have attracted thousands of scientists each year from all over Europe. It would have given a huge boost to research over here."

Marieke Degen was recently a Munich-based intern for *Nature*.

Matthias Schaefer: has good equipment but little money to run it.



The Max Planck Institute of Molecular Cell Biology and Genetics, Dresden.